

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE INFORMÁTICA**

**Departamento de Ingeniería del Software e Inteligencia Artificial**



**CORRESPONDENCIA ESTEREOSCÓPICA EN  
IMÁGENES OBTENIDAS CON PROYECCIÓN  
OMNIDIRECCIONAL PARA ENTORNOS  
FORESTALES.**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR**

**PRESENTADA POR**

**Pedro Javier Herrera Caro**

Bajo la dirección de los doctores

Gonzalo Pajares Martisanz

José Jaime Ruz Ortiz

**Madrid, 2010**

**ISBN: 978-84-693-8788-7**

**© Pedro Javier Herrera Caro, 2010**

# **CORRESPONDENCIA ESTEREOSCÓPICA EN IMÁGENES OBTENIDAS CON PROYECCIÓN OMNIDIRECCIONAL PARA ENTORNOS FORESTALES**

**TESIS DOCTORAL**



Pedro Javier Herrera Caro

Departamento de Ingeniería del Software e Inteligencia Artificial  
Facultad de Informática  
Universidad Complutense de Madrid

Madrid 2010



# **CORRESPONDENCIA ESTEREOSCÓPICA EN IMÁGENES OBTENIDAS CON PROYECCIÓN OMNIDIRECCIONAL PARA ENTORNOS FORESTALES**

*Memoria que presenta para optar al grado de Doctor en Informática*

**Pedro Javier Herrera Caro**

*Dirigida por los doctores*

**Gonzalo Pajares Martinsanz**

**José Jaime Ruz Ortiz**

Departamento de Ingeniería del Software e Inteligencia Artificial  
Facultad de Informática  
Universidad Complutense de Madrid

Madrid 2010





*A mis padres y hermana*



# Agradecimientos

En primer lugar querría agradecer a mi director toda su ayuda, disponibilidad, interés, motivación y paciencia conmigo para sacar adelante este proyecto en que decidimos embarcarnos.

A mi codirector sus exhaustivas revisiones, aportaciones, consejos y sugerencias realizadas sobre esta memoria. Y por su interés para continuar investigando en esta línea que tan buenos resultados me ha proporcionado.

A los doctores Fernando Montes e Isabel Cañellas, del Centro de Investigación Forestal (CIFOR) integrado en el Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), por el material fotográfico proporcionado y las explicaciones clarificadoras sobre el funcionamiento del sistema de medición.

A la Dra. María Guijarro su ayuda y orientación para estructurar este trabajo y por facilitarme el camino que inició ella hace unos años.

Al Dr. Jesús M. de la Cruz y al resto de miembros del grupo multidisciplinar ISCAR de la Universidad Complutense, del que formo parte, como equipo consolidado de investigación en diversas áreas entre las que se encuentra la Visión por Computador.

A la Consejería de Educación de la Comunidad de Madrid y al Fondo Social Europeo por el contrato que me ha permitido, y me permite, seguir investigando.

A la Comunidad Europea y al Consejo Nacional de Ciencia y Tecnología (CONACYT) en México, por la subvención recibida a través de FONCICYT 93829.

A la Universidad Complutense por su apoyo a la investigación, en este caso a través de la Facultad de Informática.

Una de las razones por las que estoy aquí se la debo a mis compañeros de la carrera: compañeros de prácticas y amigos. En especial a Carlos, Pablete, Ana, Nacho, Pablo M., Cris, Conchi, Jesús y Alvarito.

A mis antiguos compañeros de la Fundación Jiménez Díaz donde trabajé cuatro años. Gracias por facilitar que pudiera compaginar mis estudios con aquel trabajo que me ayudó a crecer como persona.

A mis amigos de toda la vida por estar siempre ahí.

Y finalmente, a mis padres, hermana y familia más cercana, por aguantarme todos estos años y por creer en mí.

# Resumen

Los sistemas de visión estereoscópica se han venido utilizando de forma manual desde hace varias décadas para captar información tridimensional del entorno en diferentes aplicaciones. Con el desarrollo experimentado en los últimos años por las técnicas de procesamiento computacional de imágenes, la visión estereoscópica se viene incorporando cada vez más a sistemas automáticos de diferente naturaleza. El problema central en la automatización de un sistema de visión estereoscópica es la determinación de la correspondencia entre píxeles del par de imágenes estereoscópicas que proceden del mismo punto de la escena tridimensional.

El trabajo de investigación desarrollado en esta tesis consiste en el diseño de una estrategia global para dar solución al problema de la correspondencia estereoscópica para un tipo característico de imágenes omnidireccionales procedentes de entornos forestales. Las imágenes son obtenidas mediante un sistema óptico basado en las denominadas lentes de *ojo de pez*.

Este trabajo tiene su origen en el interés suscitado por el Centro de Investigación Forestal (CIFOR) del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) para automatizar el proceso de extracción de información mediante el dispositivo de medición con número de patente MU-200501738. El interés se centra en obtener dicha información de los troncos de los árboles a partir de imágenes estereoscópicas. Con las medidas obtenidas, los técnicos realizan inventarios forestales

que incluyen estudios sobre el volumen de madera, la densidad de árboles, la evolución o crecimiento de éstos, entre otros.

La contribución principal de este trabajo consiste en la propuesta de una estrategia que combina los dos procesos esenciales en visión estereoscópica artificial como son la segmentación y correspondencia de ciertas estructuras existentes en las dos imágenes del par estereoscópico.

La estrategia se diseña para dos tipos de imágenes procedentes de sendos entornos forestales. El primero de dichos entornos se refiere a pinares de pino silvestre (*Pinus sylvestris* L.) donde las imágenes se han obtenido en días soleados y por tanto con una alta variabilidad de los niveles de intensidad debido a las zonas iluminadas. En el segundo entorno las imágenes proceden de bosques de roble rebollo (*Quercus pyrenaica* Willd.) cuya característica más relevante es que se obtienen bajo unas condiciones de iluminación relativamente escasas, días nublados o al amanecer o atardecer, pero suficiente como para producir alto contraste entre los troncos y el cielo. Debido a las características tan diferentes de ambos entornos, tanto en lo relativo a la iluminación como a la naturaleza de los propios árboles y las texturas que les rodean, los procesos de segmentación y correspondencia se diseñan atendiendo al tipo concreto de entorno forestal. Hecho éste, que marca la tendencia de la futura investigación cuando se analicen otros entornos forestales.

En el caso de los bosques de pino, el proceso de segmentación se plantea desde el punto de vista del aislamiento de los troncos mediante la exclusión de las texturas que les rodean (hojas de los pinos, suelo, cielo). Por ello, se proponen técnicas específicas de identificación de texturas para las hojas y de clasificación para el resto. En este último caso se combinan dos técnicas de clasificación clásicas como son el método de *Agrupamiento Borroso* y el *estimador paramétrico Bayesiano*. El proceso de correspondencia se plantea en dos fases. En primer lugar se identifican los píxeles homólogos en sendas imágenes del par estereoscópico mediante la adaptación a este problema de las siguientes técnicas procedentes de la teoría general de la decisión: *Integral Fuzzy de Choquet*, *Integral Fuzzy de Sugeno*, *Teoría Dempster-Shafer* y *Toma de Decisiones Multicriterio Fuzzy*. En segundo lugar, los resultados relativos a la correspondencia obtenidos mediante esas técnicas se procesan para conseguir

mejorarlos mediante la adaptación de sendos paradigmas: los *Mapas Cognitivos Fuzzy* y la *Red Neuronal de Hopfield*.

Para el segundo entorno de bosques de roble, la segmentación se plantea como un proceso de identificación de los troncos de los árboles utilizando técnicas específicas de procesamiento de imágenes, en concreto técnicas de extracción y etiquetado de regiones. Para cada región se obtiene un conjunto de atributos o propiedades que la caracterizan, y el proceso de correspondencia establece las regiones homólogas de las dos imágenes del par estereoscópico mediante medidas de similitud entre los atributos de las regiones.

La estrategia propuesta, basada en los procesos de segmentación y correspondencia, se compara favorablemente desde la perspectiva de la automatización del proceso y se plantea para su aplicación a cualquier tipo de entorno forestal, si bien con las pertinentes adaptaciones y modificaciones inherentes a los procesos de segmentación y correspondencia en función de la naturaleza del entorno forestal analizado.

### **Palabras clave**

Imágenes forestales omnidireccionales, lentes de ojo de pez, segmentación de imágenes, identificación de texturas, clasificación, crecimiento de regiones, correspondencia estereoscópica, toma de decisiones fuzzy, mapa de disparidad, filtrado del mapa de disparidad, similitud de características.





# Abstract

Stereoscopic vision systems have been used manually for decades to capture three-dimensional information of the environment in different applications. With the growth experienced in recent years by the techniques of computer image processing, stereoscopic vision has been increasingly incorporating automated systems of different nature. The central problem in the automation of a stereoscopic vision system is the determination of the correspondence between pixels of the pair of stereoscopic images that come from the same point in three-dimensional scene.

The research undertaken in this thesis comprises the design of a global strategy to solve the stereoscopic correspondence problem for a specific kind of omnidirectional image from forest environments. The images are obtained through an optical system based on the lens known as fisheye.

This work stems from the interest generated by the Forest Research Centre (CIFOR) part of the National Institute for Agriculture and Food Research and Technology (INIA) to automate the process of extracting information through the measurement mechanism with patent number MU-200501738. The focus is on obtaining this information from tree trunks using stereoscopic images. The technicians carry out forest inventories which include studies on wood volume and tree density as well as the evolution and growth of the trees with the measurements obtained.

This paper's main contribution is the proposal for a strategy that combines the two essential processes involved in artificial stereo vision: segmentation and correspondence of certain structures in the dual images of the stereoscopic pair.

The strategy is designed for two types of images from two forest environments. The first of these refers to Scots pine forests (*Pinus sylvestris* L.) where images were obtained on sunny days and therefore exhibit highly variable intensity levels due to the illuminated areas. In the second of these, the images come from Rebollo oak forests (*Quercus pyrenaica* Willd.), the main characteristic of which is that they are obtained under relatively low light conditions, on cloudy days or at dawn or dusk, but with sufficient light to produce high contrast between the trees and sky. Due to the very different characteristics of each environment - both in terms of light and the nature of trees themselves and textures that surround them - the segmentation and correspondence processes are designed specifically according to the specific type of forest environment. This sets the trend for future research when analyzing other forest environments.

In the case of pine forests, the segmentation process is approached from the point of view of isolating the trunks by excluding the textures that surround them (pine needles, the ground, the sky). For this reason, we propose the use of the specific techniques of texture identification for the pine needles and of classification for the rest. The latter case combines two classic classification techniques: *Fuzzy Clustering* and the *Bayesian Parametric estimator*. The matching process is set out in two phases. The first identifies the homogeneous pixels in separate stereo pair images, by means of the adaptation of the following techniques from general decision theory to this problem: *Choquet's Fuzzy Integral*, *Sugeno's Fuzzy Integral*, *Dempster-Shafer Theory* and *Fuzzy Multicriteria Decision Making*. Second, the results relating to correspondence obtained by these techniques are enhanced through the adaptation of two separate paradigms, namely: *Fuzzy Cognitive Maps* and the *Hopfield Neural Network*.

Regarding the second type of forest analyzed, i.e. oak, the segmentation process is designed in order to identify the tree trunks by applying specific techniques in image processing, relating to the extraction and labelling of regions, so that each region is given a set of attributes or properties that characterizes it. The matching process establishes the equivalent regions from the two stereo pair images using measures of similarity among the attributes of the regions.

The proposed strategy based on segmentation and correspondence processes can be favourably compared from the perspective of the automation of the process and we suggest it can be applied to any type of forest environment, with the appropriate adaptations inherent to the segmentation and correspondence processes in accordance with the nature of the forest environment analyzed.

**Keywords**

Forest omnidirectional images, fisheye lenses, image segmentation, texture identification, classification, region growing, stereoscopic correspondence, fuzzy decision making, disparity map, filtering disparity map, similarity of characteristics.



# Índice

<b>CAPÍTULO 1. INTRODUCCIÓN.....</b>	<b>1</b>
1.1 Antecedentes.....	1
1.2 Descripción del problema.....	6
1.2.1 Imágenes con alta variabilidad espectral	8
1.2.2 Imágenes captadas en condiciones de baja iluminación	12
1.3 Motivación y objetivos.....	16
1.3.1 Motivación	16
1.3.2 Objetivos	18
1.4 Metodología.....	19
1.5 Aportaciones de la investigación.....	20
1.6 Organización de la memoria de tesis.....	23
 <b>CAPÍTULO 2. ESTADO DEL ARTE.....</b>	 <b>25</b>
2.1 Visión general.....	25
2.2 Revisión de técnicas en clasificación de texturas.....	30
2.2.1 Metodologías de clasificación de texturas en imágenes naturales	30
2.2.2 Características de las imágenes en la clasificación de texturas naturales	35
2.2.2.1 Características espectrales a nivel de píxel	36
2.2.2.2 Características con patrones espaciales: regiones	38

2.2.2.3	Combinación de características espectrales y espaciales	42
2.2.3	Metodología de clasificación de texturas mediante información espacial	43
2.2.3.1	Información espacial mediante Campos Aleatorios de Markov	44
2.2.3.2	Relaciones estructurales	48
2.2.3.3	Otros enfoques	48
<b>2.3</b>	<b>Correspondencia estereoscópica.....</b>	<b>50</b>
2.3.1	Restricciones aplicadas en correspondencia estéreo	50
2.3.2	Técnicas en correspondencia estereoscópica	52
<b>2.4</b>	<b>Resumen y conclusiones del análisis bibliográfico.....</b>	<b>57</b>
2.4.1	Proceso completo basado en píxeles	58
2.4.2	Proceso completo basado en regiones	61
2.4.3	Resumen global	61
 <b>CAPÍTULO 3. SEGMENTACIÓN.....</b>		<b>63</b>
<b>3.1</b>	<b>Introducción.....</b>	<b>63</b>
<b>3.2</b>	<b>Segmentación basada en píxeles.....</b>	<b>65</b>
3.2.1	Generalidades	65
3.2.2	Identificación de texturas con alto contraste: Varianza	70
3.2.3	Identificación de texturas homogéneas: Clasificación	73
3.2.3.1	Planteamiento del enfoque	73
3.2.3.2	Fase de entrenamiento o aprendizaje	75
3.2.3.3	Fase de decisión	78
3.2.4	Extracción de atributos	81
3.2.4.1	Correlación	82
3.2.4.2	Color	83
3.2.4.3	Textura	83
3.2.4.4	Módulo y dirección del gradiente	84
3.2.4.5	Laplaciana	85
3.2.5	Resumen del proceso de segmentación basado en píxeles	85
<b>3.3</b>	<b>Segmentación basada en regiones.....</b>	<b>86</b>
3.3.1	Generalidades	86
3.3.2	Extracción de características	88
3.3.3	Extracción de atributos	95
3.3.4	Resumen del proceso de segmentación basado en regiones	97

<b>CAPÍTULO 4. CORRESPONDENCIA ESTEREOSCÓPICA.....</b>	<b>99</b>
4.1 Introducción.....	99
4.2 Correspondencia basada en píxeles.....	103
4.2.1 Restricciones estereoscópicas: diseño del proceso	103
4.2.2 Epipolar: geometría del sistema	109
4.2.3 Semejanza: atributos basados en el área y en las características	112
4.2.4 Unicidad especial: métodos basados en toma de decisiones	114
4.2.4.1 Integral Fuzzy de Choquet (CFI)	118
4.2.4.2 Integral Fuzzy de Sugeno (SFI)	122
4.2.4.3 Teoría de Dempster-Shafer (DES)	122
4.2.4.4 Toma de Decisiones Multicriterio Fuzzy (MCDMF)	124
4.2.5 Mapa de disparidad inicial	127
4.2.6 Continuidad de la disparidad: filtrado del mapa de disparidad	128
4.2.6.1 Mapas Cognitivos Fuzzy (FCMs)	130
4.2.6.2 Red Neuronal de Hopfield (HNN)	134
4.2.7 Resumen de parámetros en el proceso basado en píxeles	145
4.3 Correspondencia basada en regiones.....	146
4.3.1 Diseño del proceso	146
4.3.2 Epipolar: centroide	147
4.3.3 Semejanza: áreas y momentos invariantes de Hu	149
4.3.4 Orden posicional: ángulos	152
4.3.5 Resumen del proceso	153
 <b>CAPÍTULO 5. ANÁLISIS DE RESULTADOS.....</b>	 <b>157</b>
5.1 Introducción.....	157
5.1.1 Características y operación del sistema de medición estereoscópico	158
5.1.2 Enfoque basado en píxeles	163
5.1.3 Enfoque basado en regiones	168
5.2 Resultados de la Segmentación basada en píxeles.....	170
5.2.1 Agrupamiento Borroso (FC)	170
5.2.2 Estimador Paramétrico Bayesiano (PB)	173
5.2.3 Combinación Agrupamiento Borroso y Estimador Paramétrico Bayesiano	174
5.3 Resultados de la Segmentación basada en regiones.....	176



<b>5.4</b>	<b>Resultados de la Correspondencia basada en píxeles.....</b>	<b>178</b>
5.4.1	Integral Fuzzy de Choquet (CFI)	178
5.4.1.1	Cálculo de la relevancia para cada atributo	178
5.4.1.2	Resultados individuales por atributo y combinados mediante el método CFI	180
5.4.2	Integral Fuzzy de Sugeno (SFI)	184
5.4.3	Teoría de Dempster-Shafer (DES)	185
5.4.3.1	Cálculo de los vectores de medias	185
5.4.3.2	Resultados obtenidos mediante el método DES	187
5.4.4	Toma de Decisiones Multicriterio Fuzzy (MCDMF)	187
5.4.4.1	Cálculo de los pesos asociados a cada criterio	187
5.4.4.2	Resultados obtenidos mediante el método MCDMF	188
<b>5.5</b>	<b>Resultados del filtrado del Mapa de Disparidad.....</b>	<b>191</b>
5.5.1	Mapas Cognitivos Fuzzy (FCMs)	191
5.5.2	Red Neuronal de Hopfield (HNN)	196
<b>5.6</b>	<b>Métodos basados en píxeles: resumen de resultados.....</b>	<b>200</b>
<b>5.7</b>	<b>Resultados de la Correspondencia basada en regiones.....</b>	<b>202</b>
<b>5.8</b>	<b>Resumen de los resultados obtenidos basados en regiones.....</b>	<b>205</b>

<b>CAPÍTULO 6. CONCLUSIONES Y TRABAJO FUTURO.....</b>	<b>207</b>
6.1    Introducción.....	207
6.2    Conclusiones.....	207
6.3    Trabajo futuro.....	211
6.3.1    Sistema de medición estereoscópico	211
6.3.2    Procedimiento basado en píxeles	212
6.3.3    Procedimiento basado en regiones	215

<b>APÉNDICE A. SISTEMAS DE VISIÓN ESTEREOSCÓPICA.....</b>	<b>217</b>
A.1   Introducción.....	217
A.2   Visión estereográfica y señales de profundidad.....	218
A.3   Creación y visualización de imágenes estereoscópicas.....	220
A.4   Áreas de aplicación de los sistemas de visión estéreo.....	223
A.4.1   Representación de información gráfica compleja	224

A.4.2	Telepresencia	224
A.4.3	Realidad virtual	225
A.4.4	Entrenamiento de la percepción espacial	225

<b>APÉNDICE B. SEGMENTACIÓN Y CORRESPONDENCIA:</b>	
<b>NUEVOS RETOS.....</b>	<b>227</b>
B.1 Ampliación de la investigación: líneas de actuación.....	227
B.2 Resumen del estudio.....	230
<b>BIBLIOGRAFÍA.....</b>	<b>233</b>



# Lista de Figuras

Figura 1.1	(a) y (b) Imágenes estereoscópicas obtenidas mediante el dispositivo estereoscópico de medición MU-200501738.	3
Figura 1.2	Fotogrametría convencional.	4
Figura 1.3	Par de imágenes estereoscópicas omnidireccionales: (a) izquierda; (b) derecha.	11
Figura 1.4	Procesamiento radial de las imágenes.	11
Figura 1.5	Imágenes originales omnidireccionales de un par estereoscópico captadas con baja iluminación (a) Imagen izquierda; (b) Imagen derecha.	12
Figura 1.6	Imágenes omnidireccionales del par estéreo en la figura 1.5 realzadas por igualación de histograma uniforme (a) Imagen izquierda. (b) Imagen derecha.	13
Figura 2.1	Esquema del proceso de visión estereoscópica.	28
Figura 2.2	Diferentes ejemplos de texturas procedentes de las bases de imágenes Brodatz y VisTex.	39
Figura 2.3	Ejemplo de imagen omnidireccional con diferentes texturas naturales.	41
Figura 2.4	Relación de vecindad en un Campo Aleatorio de Markov.	45
Figura 3.1	Esquema del proceso de segmentación y correspondencia basado en píxeles.	69
Figura 3.2	Esquema del proceso de segmentación y correspondencia basado en regiones.	87
Figura 3.3	Perfiles de intensidad. (a) La circunferencia atraviesa una región donde los troncos y el suelo no pueden distinguirse. (b) Niveles de	91

	intensidad bajos y altos. El primero se asocia a troncos y el segundo al cielo. Las intensidades varían de 0 a 1.	
Figura 3.4	(a) Circunferencias dibujadas en amarillo y rojo. (b) Regiones homogéneas dibujadas en azul y las semillas en rojo. (c) Imagen resultante obtenida con el proceso de <i>crecimiento de regiones</i> . (d) Imagen resultante obtenida mediante el proceso de <i>etiquetado</i> .	92
Figura 3.5	(a) Semillas clasificadas dependiendo de la región a la cual pertenecen. Por claridad podemos ver las semillas pintadas con el mismo color. (b) Imagen resultante obtenida con los pasos: <i>predicción y corrección</i> . Las semillas se pintan en rojo. Cada región está delimitada con puntos amarillos teniendo en cuenta las circunferencias amarillas asociadas. (c) Imagen resultante obtenida con el segundo proceso de <i>crecimiento de regiones</i> . (d) Imagen resultante obtenida después del proceso de <i>re-etiquetado</i> .	95
Figura 4.1	Restricción de orden posicional: (a) sistema estereoscópico convencional basado en proyección de perspectiva con los ejes ópticos paralelos; (b) sistema basado en lentes de ojo de pez con ejes ópticos paralelos.	102
Figura 4.2	Opciones en la aplicación de restricciones para la obtención del mapa de disparidad.	105
Figura 4.3	Proyecciones geométricas y relaciones para el sistema de visión estéreo basado en lentes de ojo de pez.	109
Figura 4.4	Líneas epipolares en la imagen derecha generadas a partir de las posiciones en la imagen izquierda marcadas con los cuadrados.	111
Figura 4.5	Par de imágenes estereoscópicas omnidireccionales: (a) izquierda, orientada hacia el Sur geográfico; (b) derecha, orientada hacia el Norte geográfico.	115
Figura 4.6	Orientación de las cámaras en la escena 3D.	116
Figura 4.7	Esquema a considerar para la aplicación de la restricción de unicidad especial; (a) imagen izquierda; (b) imagen derecha.	116
Figura 4.8	Modelo de la Red Neuronal de Hopfield.	135
Figura 4.9	(a) Imagen izquierda señalada mostrando seis cuadrados representativos de otros tantos píxeles. (b) Líneas epipolares en la imagen derecha generadas a partir de las localizaciones en la imagen izquierda señaladas por los cuadrados.	149
Figura 5.1	Lente de ojo de pez Nikon® FC-E8.	158
Figura 5.2	Descripción del FOV de la lente Nikon® FC-E8.	159
Figura 5.3	Vista de la lente Nikon® FC-E8 acoplada sobre la cámara Nikon® Coolpix® 4500.	160
Figura 5.4	Vista del sistema de medición estereoscópico general.	161
Figura 5.5	Aislamiento de los troncos de interés.	165
Figura 5.6	Etiquetado de los troncos.	166
Figura 5.7	Imágenes izquierda y derecha segmentadas utilizando el descriptor de texturas junto con el clasificador FC.	172

Figura 5.8	Imágenes izquierda y derecha segmentadas utilizando el descriptor de texturas junto con el clasificador <i>PB</i> .	174
Figura 5.9	Imagen segmentada mediante la combinación de los clasificadores <i>FC</i> y <i>PB</i> .	175
Figura 5.10	Etiquetado de regiones: (a) imagen izquierda, (b) imagen derecha. Cada región aparece identificada por un único número.	176
Figura 5.11	Porcentajes promediados de error obtenidos para cada atributo individual a partir del conjunto de pares de imágenes estereoscópicas de entrenamiento.	180
Figura 5.12	Relevancias individuales obtenidas para cada atributo individual a partir del conjunto de pares de imágenes estereoscópicas de entrenamiento.	180
Figura 5.13	Mapa de disparidad inicial obtenido con la propuesta <i>CFI</i> : (a) sin segmentación; (b) con segmentación según el proceso combinado.	181
Figura 5.14	Porcentajes promediados de error obtenidos por aplicación de los criterios por separado y en combinación a través de los métodos <i>YAG</i> y <i>CFI</i> .	184
Figura 5.15	Desviaciones estándar promediadas obtenidas por aplicación de los criterios por separado y en combinación a través de los métodos <i>YAG</i> y <i>CFI</i> .	184
Figura 5.16	Vectores de medias obtenidos para las clases de correspondencias positivas y negativas a partir del conjunto de pares de imágenes de entrenamiento.	186
Figura 5.17	Mapa de disparidad inicial obtenido con la propuesta <i>MCDMF</i> .	189
Figura 5.18	(a) Porcentajes promediados y (b) desviaciones estándar obtenidas por los métodos <i>CFI</i> , <i>SFI</i> , <i>DES</i> y <i>MCDMF</i> .	190
Figura 5.19	Mapa de disparidad obtenido con la propuesta <i>FCM</i> .	192
Figura 5.20	(a) Imagen izquierda omnidireccional; (b) área expandida correspondiente al área marcada en la imagen de la figura 5.20(a); (c) mapa de disparidad obtenido con la propuesta <i>MCDMF</i> ; (d) mapa de disparidad obtenido con la propuesta <i>FCM</i> .	193
Figura 5.21	Porcentajes promediados de error obtenidos por el enfoque <i>FCM</i> frente a los métodos <i>CFI</i> , <i>SFI</i> , <i>DES</i> y <i>MCDMF</i> .	195
Figura 5.22	Desviaciones estándar obtenidas por el enfoque <i>FCM</i> frente a los métodos <i>CFI</i> , <i>SFI</i> , <i>DES</i> y <i>MCDMF</i> .	195
Figura 5.23	Mapa de disparidad obtenido con la propuesta <i>HNN</i> .	196
Figura 5.24	(a) Imagen izquierda omnidireccional; (b) área expandida correspondiente al área señalada en la imagen de la figura 5.24(a); (c) mapa de disparidad obtenido mediante la propuesta <i>MCDMF</i> . (d) mapa de disparidad obtenido con la propuesta <i>HNN</i> .	197
Figura 5.25	(a) Variación promedio de la función de energía con el número de iteraciones; (b) variación del porcentaje promediado de error frente al número de iteraciones.	199
Figura 5.26	Porcentajes promediados de error obtenidos por el enfoque <i>HNN</i> frente a cada una de las propuestas de toma de decisiones objeto de investigación: <i>CFI</i> , <i>SFI</i> , <i>DES</i> y <i>MCDMF</i> .	200

Figura 5.27	Desviaciones estándar obtenidas por el enfoque <i>HNN</i> frente a cada una de las propuestas de toma de decisiones objeto de investigación: <i>CFI</i> , <i>SFI</i> , <i>DES</i> y <i>MCDMF</i> .	200
Figura 5.28	Propuesta de proceso de correspondencia estereoscópica basada en píxeles.	202
Figura 5.29	Propuesta de proceso de correspondencia estereoscópica basada en regiones.	206
Figura 6.1	Par estereoscópico generado con Truevision3D (a) imagen izquierda; (b) imagen derecha.	212
Figura 6.2	Robot equipado con sistema de visión estereoscópica convencional para navegación autónoma.	215
Figura A.1	(a) Sistema de visión estereoscópica biológico, con dos objetos en la escena 3D (estrella y triángulo) a distintos niveles de profundidad; (b) Superposición de las imágenes de ambos ojos, donde se observa que la separación relativa entre los triángulos es menor que entre las estrellas, lo que está en relación inversa con la distancia de los correspondientes objetos en la escena 3D.	218
Figura A.2	Señales de profundidad: perspectiva, solapamiento, desviación cromática y sombreado.	219
Figura A.3	Visor estereoscópico de finales del siglo XIX.	220
Figura A.4	Anaglifos correspondientes a dos pares estereoscópicos analizados.	221
Figura A.5	Gafas para visualización de Anaglifos.	222
Figura A.6	Ejemplo del polarizador de la empresa Dimap.	222
Figura A.7	Ejemplos de sistemas para visualización estereoscópica: (a) lentes polarizadas; (b) gafas obturadoras.	223
Figura A.8	Visualización de fotografías aéreas estereoscópicas en Fotogeología, mediante visión paralela asistida con lentes.	224
Figura B.1	Par de imágenes estereoscópicas omnidireccionales procedentes de robles: (a) izquierda; (b) derecha.	228
Figura B.2	Imágenes realzadas a partir de las mostradas en la figura B1 mediante igualación de histograma sobre la componente de intensidad en el modelo de color <i>HSI</i> .	229
Figura B.3	(a) y (b) Recortes homólogos procedentes de las imágenes estereoscópicas B.2(a) y (b) respectivamente; (c) resultado del proceso de segmentación considerando sólo texturas de cielo y hojas junto con hierba del suelo; (d) disparidades obtenidas con el rango de variación mostrado en la barra de color.	229
Figura B.4	(a) Resultado de la segmentación manual; (b) disparidades obtenidas con la variación mostrada en la barra de color.	230
Figura B.5	(a) Imagen de la figura B.1(a) en el modelo de color <i>HSI</i> ; (b) segmentación mediante la aplicación del método de Otsu.	231

## Lista de Tablas

Tabla 2.1	Ventajas e inconvenientes de los diferentes modelos de color.	37
Tabla 3.1	Parámetros y umbrales implicados en el proceso de segmentación basado en píxeles.	86
Tabla 3.2	Umbrales implicados en el proceso de segmentación basado en regiones.	97
Tabla 4.1	Comportamiento del término energía frente a los datos y las consistencias contextuales.	142
Tabla 4.2	Parámetros y umbrales implicados en el proceso de correspondencia basado en píxeles.	146
Tabla 4.3	Parámetros y umbrales implicados en el proceso de correspondencia basado en regiones.	155
Tabla 5.1	Disparidades promediadas para el <i>ground truth</i> .	166
Tabla 5.2	Centros de las clases para las texturas hierba y cielo.	171
Tabla 5.3	Centros de las clases para las texturas hierba y cielo obtenidas mediante <i>PB</i> .	173
Tabla 5.4	Matrices de covarianza de las clases para las texturas hierba y cielo obtenidas mediante <i>PB</i> .	173
Tabla 5.5	Porcentajes promediados de error y relevancias individuales obtenidas para cada atributo con el conjunto de pares de imágenes de entrenamiento.	179
Tabla 5.6	Porcentajes promediados de error y desviaciones estándar según los seis atributos individuales, el método combinado de Yager y la propuesta basada en la toma de decisiones <i>CFI</i> .	183



Tabla 5.7	Vectores de medias $\bar{v}_j$ obtenidos para las clases de correspondencias positivas ( $w_1$ ) y negativas ( $w_2$ ) a partir del conjunto de pares de imágenes de entrenamiento.	186
Tabla 5.8	Porcentajes promediados de error y pesos específicos obtenidos para cada atributo con el conjunto de pares de imágenes de entrenamiento.	188
Tabla 5.9	Porcentajes promediados de error y desviaciones estándar para los métodos: <i>CFI</i> , <i>SFI</i> , <i>DES</i> y <i>MCDMF</i> .	189
Tabla 5.10	Porcentajes promediados de error y desviaciones estándar obtenidas por el enfoque <i>FCM</i> frente a las propuestas de toma de decisiones objeto de investigación.	195
Tabla 5.11	Porcentajes promediados de error y desviaciones estándar obtenidas por el enfoque <i>HNN</i> frente a las propuestas de toma de decisiones objeto de investigación.	199
Tabla 5.12	Porcentajes promediados de error y desviaciones estándar obtenidas para cada atributo por separado y para los métodos <i>YAG</i> , <i>CFI</i> , <i>SFI</i> , <i>DES</i> , <i>MCDMF</i> , <i>FCM</i> y <i>HNN</i> , resaltando en negrita los mejores resultados en cada categoría.	201
Tabla 5.13	Resultados obtenidos empleando los momentos de Hu para establecer la correspondencia de regiones en ambos pares estereoscópicos.	203

# Lista de Abreviaturas

ACM	<i>Active Contour Model</i> , Contorno Activo.
CCD	<i>Charge Coupled Device</i> , Dispositivo de Carga Acoplada.
CFI	<i>Choquet Fuzzy Integral</i> , Integral Fuzzy de Choquet.
CICYT	Centro de Investigación Científica y Tecnológica.
CIE	<i>Commission Internationale de l'Éclairage</i> , Comité Internacional de Iluminación.
CIFOR	Centro de Investigación Forestal.
CMYK	C, cyan; M, magenta; Y, <i>yellow</i> -amarillo; K, <i>black</i> -negro.
CSIC	Consejo Superior de Investigaciones Científicas.
DEM	<i>Digital Elevation Models</i> , Mapas Digitales del Terreno.
DES	<i>Dempster-Shafer theory</i> , Teoría de Dempster-Shafer.
EM	<i>Expectation Maximization</i> , Máxima Expectación.
FC	<i>Fuzzy Clustering</i> , Agrupamiento Borroso.
FCM	<i>Fuzzy Cognitive Map</i> , Mapa Cognitivo Fuzzy.
FOV	<i>Field Of View</i> , Campo de vision.
GPS	<i>Global Positioning System</i> , Sistema de Posicionamiento Global.
HNN	<i>Hopfield Neural Network</i> , Red Neuronal de Hopfield.
HSI	H, <i>hue</i> -matiz; S, <i>saturation</i> -saturación; I, <i>intensity</i> -intensidad.
INIA	Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria.
INTA	Instituto Nacional de Técnica Aeroespacial.
ISCAR	Ingeniería de Sistemas, Control, Automatización y Robótica.

KNN	<i>k-Nearest Neighbours</i> , vecinos más próximos.
LOU	Ley Orgánica de Universidades.
LVQ	<i>Learning Vector Quantization</i> , Cuantización Vectorial.
MCDM	<i>Multi-Criteria Decision Making</i> , Toma de Decisiones Multicriterio.
MCDMF	<i>Multi-Criteria Decision Making Fuzzy</i> , Teoría de la Decisión Multicriterio Fuzzy, Toma de Decisiones Multicriterio Fuzzy.
MRF	<i>Markov Random Field</i> , Campo Aleatorio de Markov.
PB	Estimador Paramétrico Bayesiano.
RGB	R, <i>red</i> -rojo; G, <i>green</i> -verde; B, <i>blue</i> -azul.
SFI	<i>Sugeno Fuzzy Integral</i> , Integral Fuzzy de Sugeno.
SOM	<i>Self Organizing feature Maps</i> , Mapas Auto-Organizativos.
YIQ	I, <i>in-fase</i> -en fase; Q, <i>quadrature</i> -cuadratura.

# 1

## Introducción

### 1.1 Antecedentes

Los inventarios forestales proporcionan información sobre la que se basa la gestión de los bosques. En el inventario se hacen mediciones con el fin de determinar diversos parámetros sobre la estructura poblacional y el estado de la masa forestal y cuantificar las existencias de madera, fruto, resina u otros productos forestales. Los primeros inventarios forestales consistían en un conteo de todos los árboles, clasificándolos en función de su clase diamétrica (los diámetros de los árboles se clasifican en unos intervalos de amplitud fija, normalmente 5 o 10cm), denominándose esta técnica *inventario por conteo pie a pie* (Pardé y Bouchon, 1987). A mediados del siglo XX, conforme iba encareciéndose la mano de obra, empezó a generalizarse el uso de los inventarios por muestreo, destacando en España el inicio del 1<sup>er</sup> inventario Forestal Nacional en 1966. En estos inventarios se miden los árboles en parcelas de muestreo cuyo tamaño y número depende de las características de la masa forestal y de la precisión requerida (Pita, 1973; Gregoire, 1998). Las parcelas suelen replantearse distribuidas de forma sistemática por toda la superficie arbolada, y en ellas se miden los diámetros de los árboles a la altura de 1,30m (denominados *diámetros normales*) de todos los pies incluidos dentro de la parcela, midiéndose además en una submuestra de árboles otras variables más costosas de medir, como la altura, diámetros de copa, etc. (Mandallaz y Ye, 1999). A partir de esta submuestra de árboles se ajustan ecuaciones

precisas de volumen de madera u otras variables de interés en función del diámetro normal, dichas ecuaciones se aplican al resto de árboles medidos en la parcela para obtener los valores por parcela, a partir de los cuales mediante técnicas de estimación tradicionales basadas en el diseño de muestreo aleatorio o mediante técnicas geoestadísticas desarrolladas en las últimas décadas se estima la densidad, normalmente mediante el área basimétrica (superficie de la sección de los troncos de los árboles a la altura de 1,30m referida a una hectárea) o la producción (Montes y col. 2005). A partir de inventarios sucesivos, que en los montes ordenados suelen realizarse cada 10 años, se conoce la evolución de la masa forestal en el tiempo y se planifican los tratamientos selvícolas a realizar. Los troncos, técnicamente denominados *fustes* cuando tienen un aprovechamiento, como pueda ser medir el volumen, de ahora en adelante nos referiremos a ellos con la primera expresión con el fin de mantenernos en un lenguaje menos específico dado el carácter computacional de la tesis.

En 2005, el Centro de Investigación Forestal español (CIFOR) integrado en el Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) patentó el dispositivo de medición de bosques MU-200501738. La empresa “Consultoría Ferroviaria y Medioambiental, S.L.” desarrolló para el INIA un prototipo del dispositivo de medición, adaptado para una cámara digital Nikon® Coolpix® 4500 con una lente de “*ojo de pez*” del tipo FC-E8. Este dispositivo, cuya descripción se detalla en el capítulo cinco, no es ni más ni menos que un sistema cuyo fundamento es la visión estereoscópica. La cámara, junto con su sistema óptico, se sitúa en dos posiciones diferentes, lo que permite obtener dos imágenes ligeramente desplazadas de la misma escena, este hecho constituye precisamente el fundamento de la visión estereoscópica.

En la figura 1.1(a) y (b) se muestra un par de imágenes estereoscópicas obtenidas mediante el dispositivo de medición mencionado previamente.

Resulta bien conocido el hecho de que un sistema estereoscópico permite determinar distancias desde el propio sistema a los objetos en la escena, ver apéndice A. A pesar de las diferencias que este sistema presenta respecto de los sistemas estereoscópicos convencionales debido al diferente sistema de proyección de la escena sobre el dispositivo de captura que genera la imagen, el dispositivo mencionado permite también la medición de distancias desde él a determinados puntos de la escena, lo que de algún modo justifica su sobrenombre en cuanto a dispositivo de medición, siendo por

otra parte su objetivo final. En efecto, tal y como se ha mencionado previamente, varios de los parámetros a medir en relación a la realización de inventarios forestales, se basan en medidas de distancias, que permitirán indirectamente calcular las alturas de los árboles, el espesor de los troncos o las propias distancias a los árboles con el fin de estimar la densidad de los mismos en la zona bajo análisis, entre otras varias.

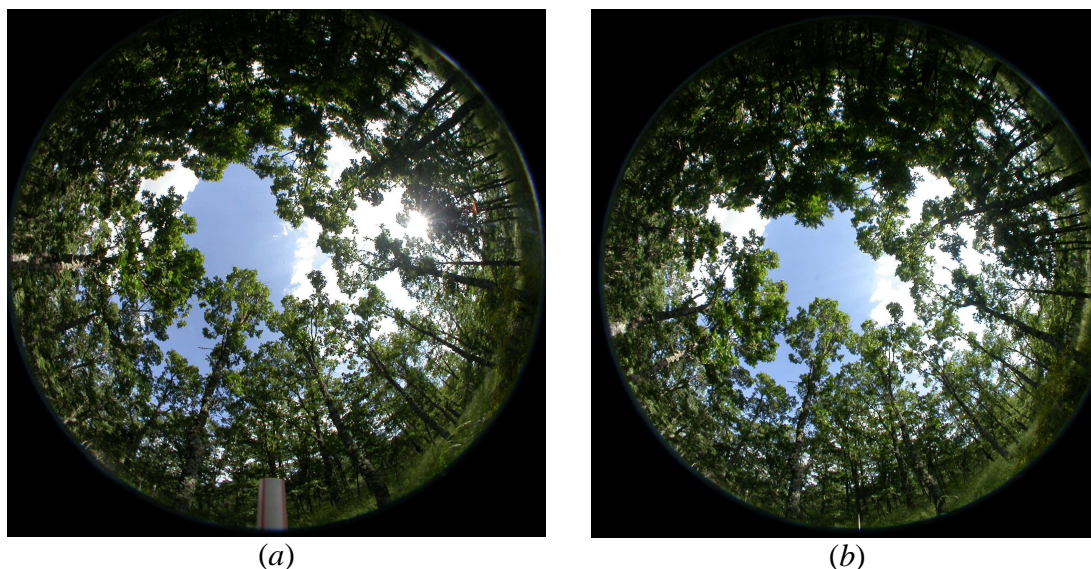


Figura 1.1 (a) y (b) Imágenes estereoscópicas obtenidas mediante el dispositivo estereoscópico de medición MU-200501738

Hasta la fecha actual, todas estas medidas se han venido realizando generalmente *in situ*, de forma que sobre las parcelas de muestreo se toman en campo las medidas pertinentes mediante los dispositivos apropiados, lo que hace que el inventario sea una tarea muy costosa, por lo que en la actualidad se están buscando técnicas, como el LIDAR o la teledetección, que consuman menos recursos y permitan tener la información en intervalos de tiempo más cortos. A partir de la utilización del dispositivo MU-200501738, por parte del equipo de investigación forestal del INIA, algunas de tales medidas se vienen realizando de forma semiautomática. Para ello, cada una de las parcelas de muestreo se encuentra perfectamente identificada y localizada mediante posicionamiento espacial a base de marcas sobre el terreno (*landmarks*) y ayuda de GPS (*Global Position System*). Cada vez que se quiere realizar una medición sobre dicha parcela, con el fin de realizar cualquier análisis comparativo, respecto de mediciones previas, se fija siempre el dispositivo en la misma localización espacial. Además, su posicionamiento relativo debe ser siempre el mismo, de este modo, una imagen, digamos la izquierda del par estereoscópico, siempre se obtiene con la misma orientación espacial del dispositivo, con la cámara situada en el extremo orientado hacia

el Sur geográfico de la guía de desplazamiento; la segunda imagen o imagen derecha del par se obtiene con la cámara situada en el extremo Norte de la guía, tras su desplazamiento de Sur a Norte, tal y como se describe en el capítulo cinco. Nada impide que la orientación sea en relación a cualquiera de las coordenadas geográficas, eso sí siempre que se quiera realizar una medida comparativa, en diferentes instantes de tiempo, debe ser siempre la misma. El eje óptico de la lente se mantiene paralelo antes y después del desplazamiento de la cámara, por lo que las dos imágenes capturadas constituyen lo que se denomina un *par estereoscópico*, de suerte que conociendo el desplazamiento relativo entre dos puntos del par de imágenes estereoscópicas, que representan la misma entidad física tridimensional, por ejemplo un determinado punto a cierta altura en el tronco del árbol, se puede determinar la posición relativa de dicho punto en el espacio 3D respecto del dispositivo de medida.

Un grado de avance respecto de los procedimientos de medida realizados directamente en campo en las parcelas de muestreo lo constituye el hecho de disponer de un dispositivo de captura de imágenes estereoscópicas, que convenientemente impresas sobre papel permite al estereógrafo determinar manualmente los mencionados desplazamientos relativos. En la figura 1.2 se muestra una imagen relativa al proceso fotogramétrico mencionado.



Figura 1.2 Fotogrametría convencional

Actualmente, el equipo de investigación del INIA determina dichos desplazamientos relativos entre puntos homólogos del par estereoscópico mediante un programa informático basado en el trazado de líneas y puntos sobre la pantalla del computador. Esta tarea requiere, naturalmente, la presencia de un operador, encargado

de determinar qué punto de una imagen representa la misma entidad física en la otra imagen del par estereoscópico. Una vez establecidos estos desplazamientos relativos, conocidos técnicamente como *disparidades*, es posible determinar las deseadas distancias desde el dispositivo a las entidades del mundo 3D que han originado los desplazamientos en las imágenes. Esto se describe con detalle en el capítulo cuatro donde se establecen las relaciones geométricas del sistema de medición estereoscópica. Por el mero hecho de calcular las disparidades mediante la intervención del usuario, se dice al respecto que las medidas se realizan de forma semiautomática. Desde el punto de vista del trabajo de investigación presentado en esta tesis, el objetivo se centra precisamente en la automatización del proceso, hasta donde sea posible, de forma que las mencionadas disparidades se puedan establecer de forma automática.

Volviendo sobre los sistemas de ojo de pez, en general, éstos se caracterizan por su capacidad para captar imágenes de la escena tridimensional (3D) en un amplio campo de visión en torno a las cámaras. En el caso del sistema descrito previamente la escena 3D capturada se sitúa bajo un recorrido angular de  $183^{\circ} \times 360^{\circ}$ . En el capítulo cinco se describe esto con más detalle.

Desde el punto de vista de la aplicación del sistema para la realización de inventarios forestales, conviene destacar el hecho de que la proyección de la escena 3D en estos sistemas en forma de imagen representa una importante ventaja respecto de los sistemas convencionales basados en sistemas ópticos con proyección de perspectiva. En efecto, gracias a la óptica basada en ojo de pez, los árboles aparecen completos en las imágenes, desde la base a la copa, tanto los que se sitúan próximos al sistema en la escena 3D como los más alejados. Por el contrario, en los sistemas equipados con lentes convencionales bajo la mencionada proyección de perspectiva, los objetos cercanos se proyectan parcialmente, es decir, lo más probable es que su imagen no aparezca completa, este hecho ha sido remarcado y analizado en las investigaciones llevadas a cabo por Abraham y Förstner (2005).

Debido a que los árboles aparecen proyectados al completo en las imágenes, el sistema estereoscópico permite el cálculo de las posiciones respecto al dispositivo de los puntos significativos de los árboles en la escena 3D, como los puntos que determinan los diámetros a diferentes alturas lo largo del tronco, las alturas de los árboles, así como la posición de los árboles en la parcela de muestreo. Estos datos permiten elaborar



precisas ecuaciones de perfil de troncos, calcular los diámetros y alturas de los árboles o variables de copa. A partir de estos valores se puede estimar la densidad de árboles en un área determinada, el volumen de madera en dicha área, el área basimétrica, así como otras variables de interés (Montes y col. 2009). Además de lo anterior, la conservación de tales imágenes a lo largo del tiempo constituye un registro permanente de la zona de muestreo que permite el control de errores de medición así como la aplicación en el futuro de técnicas basadas en minería de datos, con la ventaja que representa el hecho de no tener que volver a visitar la parcela, tanto desde el punto de vista económico como ecológico. Otra de las ventajas importantes de estos sistemas, también desde el punto de vista ecológico, se refiere a su naturaleza pasiva, esto es, las imágenes se capturan sin más intervención sobre la escena que la requerida por la ubicación del sistema, lo que evita la agresión de los sistemas biológicos presentes en el bosque durante el proceso de captura de imágenes. Esta es una ventaja frente a otros sistemas activos tales como los basados en sensores láser.

## **1.2 Descripción del problema**

Centrado ya el problema sobre la conveniencia de utilizar un sistema de visión estereoscópica para la realización de medidas pasivas en entornos forestales y de este modo poder estimar los parámetros necesarios para la realización de los inventarios forestales; ahora nuestro problema se centra en el desarrollo de técnicas y estrategias automáticas para obtener el máximo rendimiento del dispositivo de medición citado previamente y llegar así a un mejor aprovechamiento y mantenimiento de los recursos forestales. En este sentido, la cuestión se orienta ahora en identificar el proceso o procesos más apropiados relativos a la visión estereoscópica para llevar a cabo hasta donde sea posible el proceso de automatización. Como se ha mencionado anteriormente, un sistema estereoscópico basado en lentes de ojo de pez permite la captura de una amplia zona del espacio forestal 3D con la ventaja de que los árboles aparecen proyectados en su totalidad en las imágenes, razón por la cual, como ya se ha mencionado previamente, estos sistemas resultan apropiados para tal propósito. No obstante, no están exentos de una cierta problemática relativa a la forma y condiciones en las que se obtienen las imágenes, que analizamos seguidamente.

Bien es cierto, que para llegar a medir distancias se necesita realizar un proceso de visión estereoscópica al completo, entendiendo por tal, la realización de todos los

pasos necesarios, tal y como se explica en capítulos sucesivos, desde la captura de las imágenes hasta la obtención de las distancias y posteriormente la medida de volúmenes, densidades, alturas u otros parámetros de interés. Debido a la complejidad de las imágenes analizadas, como se verá seguidamente, existen dos pasos intermedios dentro del proceso global, que por su complejidad intrínseca requieren por sí solos un estudio pormenorizado y especialmente dedicado. Estos pasos son exactamente el problema de la identificación de los troncos como estructuras de interés en las imágenes y su ubicación en sendas imágenes del par estereoscópico. Posteriormente, llegaremos a la descripción y tratamiento pormenorizado de estos mismos dos problemas como integrantes del proceso de visión estereoscópica. Es en ellos donde se concentra el esfuerzo de investigación reflejado en este trabajo de tesis, ya que por un lado son los pasos clave del proceso general y por otro resultan ser los de más difícil solución, encontrándose actualmente abiertos a todo tipo de propuestas en la literatura universal especializada.

Algunas características inherentes al propio dispositivo de captura o sensor y su diseño son las que se mencionan seguidamente. Al tratarse de un sistema estereoscópico, las cámaras se sitúan de forma que entre ellas existe lo que técnicamente en la literatura especializada se conoce como *línea base* (Pajares y Cruz, 2007a) o “*base-line*” en terminología inglesa. En el caso que nos ocupa, relativo al dispositivo MU-200501738 la línea base permite una separación de hasta 1m. Por otro lado, las imágenes se obtienen mediante la cámara digital anteriormente mencionada, la cual se encuentra equipada con un dispositivo CCD (*Charge Coupled Device*) con  $2272 \times 1704$  elementos sensores, siendo éste el tamaño relativo a la resolución espacial en píxeles respecto de las dimensiones ancho y alto; si bien, debido al sistema de proyección basado en las lentes de ojo de pez, sólo la zona central delimitada por el círculo que contiene 2045059 píxeles aparece con información de la escena 3D. Dicho círculo está circunscrito en un cuadrado de  $1616 \times 1616$  píxeles.

Todas las imágenes originales se adquieren en el espacio de color *RGB* (*Red-rojo*, *Green-verde*, *Blue-azul*), cuyas características se han estudiado en numerosos trabajos sobre la teoría del color (Pajares y Cruz, 2007a; Gonzalez y Woods, 2008). El estudio desarrollado en el presente trabajo se centra en el tratamiento estereoscópico de dos tipos de imágenes pertenecientes a sendas clases de bosques. Las primeras

pertenecen a bosques de pinares, en concreto a pino silvestre (*Pinus sylvestris* L.), cuya característica más relevante es que presentan una alta variabilidad de los niveles de intensidad en las tres componentes espectrales  $R$ ,  $G$ ,  $B$  por el hecho de haber sido captadas en un día soleado; mientras que las segundas proceden de bosques de robles, en concreto a roble rebollo (*Quercus pyrenaica* Willd.), habiendo sido obtenidas bajo unas condiciones de escasa iluminación en un día nublado o al atardecer o anochecer en un día soleado. De ahora en adelante nos referiremos a los pinos silvestres simplemente como pinos, mientras que a los robles rebollos como robles. A continuación se explica la problemática inherente a cada una de las clases de imágenes mencionadas. A pesar de que en ambos casos se trata de imágenes estereoscópicas, su procesamiento debe necesariamente abordarse desde distintas perspectivas. Esto va a originar, como se expondrá a lo largo de esta memoria, diferentes procedimientos relativos al proceso de visión estereoscópica. A continuación exponemos la problemática inherente a los dos tipos de imágenes mencionados.

### 1.2.1 Imágenes con alta variabilidad espectral

La figura 1.3(a) y (b) muestra un par de imágenes estereoscópicas pertenecientes al primer tipo. Son imágenes omnidireccionales obtenidas con el sistema descrito previamente y representativas del conjunto de imágenes capturadas en el área de muestreo del pinar. Conviene recordar que los troncos de los árboles son los componentes básicos desde el punto de vista del análisis forestal, ya que por un lado contienen el mayor volumen de madera del árbol y por otro identifican la posición del árbol mismo en el área bajo análisis, a partir de los cuales se pueden estimar variables tales como la densidad de árboles, alturas de los mismos u otras ya mencionadas. Resulta fácil apreciar, a partir de esas imágenes, cómo los rayos incidentes del sol producen una importante variabilidad en la iluminación en una parte de la imagen (zona superior derecha), de tal forma que los troncos en esta zona presentan tonalidades claras por su exposición directa al sol. En cambio, el resto de troncos en la imagen presentan tonalidades oscuras por no encontrarse expuestos directamente al sol, y por tanto en la sombra. Además, como las imágenes son capturadas con una cámara desde dos posiciones distintas, que difieren una cierta distancia, conocida como *línea base*, las copas de los árboles se localizan en diferentes situaciones con respecto a cada posición de la cámara; esto hace que los rayos incidentes del sol produzcan una importante

variabilidad en la iluminación en ambas imágenes, distinguiéndose áreas muy iluminadas rodeadas por otras en sombra, y viceversa, aún a pesar de representar la misma escena. Además de lo anterior, en función de la densidad de los árboles, aparecen fenómenos de oclusión, donde determinadas estructuras, en general troncos, están presentes en una imagen y ausentes en la otra del par estereoscópico. En la figura 1.3, en la parte superior de las imágenes, se pone de manifiesto este último fenómeno. Estos factores determinan que los procesos encaminados al tratamiento de estas imágenes desde el punto de vista estereoscópico resulte una tarea ciertamente complicada, ya que se trata de una problemática presente de forma general en este tipo de imágenes.

Desde el punto de vista del procesamiento estereoscópico y a la hora de tratar de identificar qué tronco de una imagen se corresponde con qué tronco de la otra imagen del par estereoscópico, uno podría pensar en utilizar descriptores de textura con el fin de identificar el mismo tronco en sendas imágenes, lo que se conoce como *correspondencia estereoscópica* y constituye por otro lado la parte esencial del proceso. Debido a las diferentes características espectrales mostradas y a la problemática exhibida por este tipo de troncos según su ubicación en el espacio y por tanto en la imagen, su identificación mediante este procedimiento no resulta apropiada, entre otras razones porque las imágenes no presentan patrones de textura homogéneos que faciliten su reconocimiento. Obsérvese en las imágenes 1.3(a) y (b) las diferentes texturas exhibidas por los troncos aún a pesar de pertenecer a la misma especie de árbol. Por ello, es necesario ahondar más en el estudio y análisis de la problemática presente en estas imágenes.

Tras un análisis más en profundidad sobre este tipo de imágenes, además de las texturas propias de los troncos, se pueden distinguir básicamente otros tres tipos de texturas, a saber:

- hierba en el suelo
- cielo entre los huecos de las copas de los árboles
- hojas de los árboles.

Desde el punto de vista del procesamiento de las imágenes y considerando que el interés se centra en la identificación de los troncos, es evidente que estos tres tipos de texturas no constituyen el objetivo del proceso de la visión estereoscópica, por lo que

nuestro interés sobre ellas se centra precisamente en su exclusión del proceso de correspondencia. Por este motivo, en un primer paso del procesamiento se trata de identificar y establecer los métodos necesarios para su exclusión, de forma que no intervengan en los procesos posteriores, facilitando de este modo lo que será el proceso de correspondencia para los troncos de los árboles. Este primer paso, se identifica y caracteriza, a partir de este momento, como *proceso de segmentación*. En el presente trabajo se propone una estrategia encaminada a la identificación de las texturas objeto de exclusión tal y como se expone en el capítulo tres. La propuesta de dicha estrategia ha requerido un estudio en profundidad sobre diversos métodos de identificación de texturas en general y en entornos naturales en particular. Además, existen otros tipos de entornos forestales, tal como robles con similar problemática en cuanto a condiciones de iluminación se refiere y con un agravamiento en lo relativo a la identificación de las texturas objeto de exclusión, ver apéndice B. Ciertamente, la complejidad del problema es tal que no ha sido posible hasta el momento obtener un resultado satisfactorio mediante segmentación, planteándolo como un trabajo a abordar en el futuro. En consecuencia, ateniéndonos tanto a la propuesta de segmentación realizada en este trabajo como a la que sin duda surgirá en el futuro, se ha considerado de vital importancia la identificación del mayor número posible de métodos de clasificación de texturas naturales, tanto por lo que resta por investigar como por las posibles mejoras que puedan surgir sobre la propuesta realizada. Estas son principalmente las razones por las que se realiza una extensa revisión de técnicas de clasificación de texturas en el capítulo dos.

Una vez identificadas y excluidas las texturas anteriores, el siguiente objetivo consiste en establecer la correspondencia entre los troncos del par de imágenes estereoscópicas. Este segundo paso recibe de ahora en adelante el nombre de *proceso de correspondencia*, al que se dedica un importante espacio en el presente trabajo.

Volviendo sobre las imágenes de la figura 1.3(a) y (b), una vez descartada la identificación de los troncos mediante análisis de texturas, podría pensarse en utilizar la distribución radial de los troncos como elemento descriptor, de forma que mediante un barrido radial se pudieran determinar líneas radiales con idénticas propiedades o características espectrales, figura 1.4. Tras diversos experimentos previos realizados en este sentido se llegó a la conclusión de que la problemática relativa a las variaciones espectrales persiste igualmente (Suárez, 2009). Además, a pesar de la apariencia casi

radial de los troncos, en la realidad no es tal; ello es debido a que o bien los propios árboles no crecen perfectamente verticales o el propio dispositivo de medida no se posiciona exactamente horizontal sobre la superficie del terreno. Ello hace que este enfoque no produzca los resultados esperados. Por todo ello, y una vez descartado el enfoque de barrido radial, así como la identificación de los troncos como posibles regiones homogéneas, se determina que las características más apropiadas para la correspondencia de los troncos en este tipo de imágenes son los píxeles, de suerte que el enfoque finalmente adoptado consiste en establecer la correspondencia entre los píxeles que no han sido excluidos en el proceso de segmentación previo y sobre los que se supone se encontrarán los propios troncos, objeto principal del proceso de correspondencia.

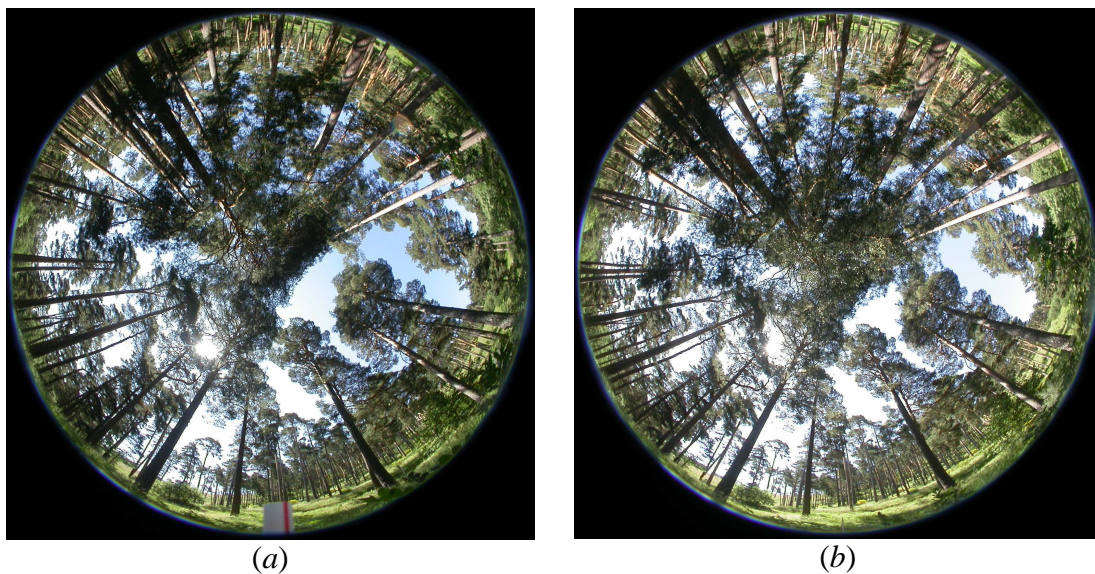


Figura 1.3 Par de imágenes estereoscópicas omnidireccionales: (a) izquierda; (b) derecha.



Figura 1.4 Procesamiento radial de las imágenes.

### 1.2.2 Imágenes captadas en condiciones de baja iluminación

La figura 1.5(a) y (b) muestra un par de imágenes estereoscópicas omnidireccionales, captadas con el mismo dispositivo de medición que en el caso anterior, si bien pertenecientes a un bosque de robles y como se ha mencionado previamente, bajo unas condiciones de escasa iluminación, en este caso como consecuencia de haberse obtenido en un día nublado. Este par de imágenes es representativo e ilustrativo del conjunto de imágenes de estas características para las que se ha diseñado una estrategia específica de segmentación y correspondencia. El requisito esencial para poder aplicar dicha estrategia es que la captura de dichas imágenes se realice bajo las mencionadas condiciones de escasa iluminación, en este sentido se pueden conseguir este tipo de imágenes además de en días nublados, durante el amanecer o anoecer de cualquier otro día, siempre y cuando los niveles de iluminación no produzcan los efectos descritos en el primer tipo de imágenes. Esto permite la ampliación del rango de operatividad del sistema de medición.

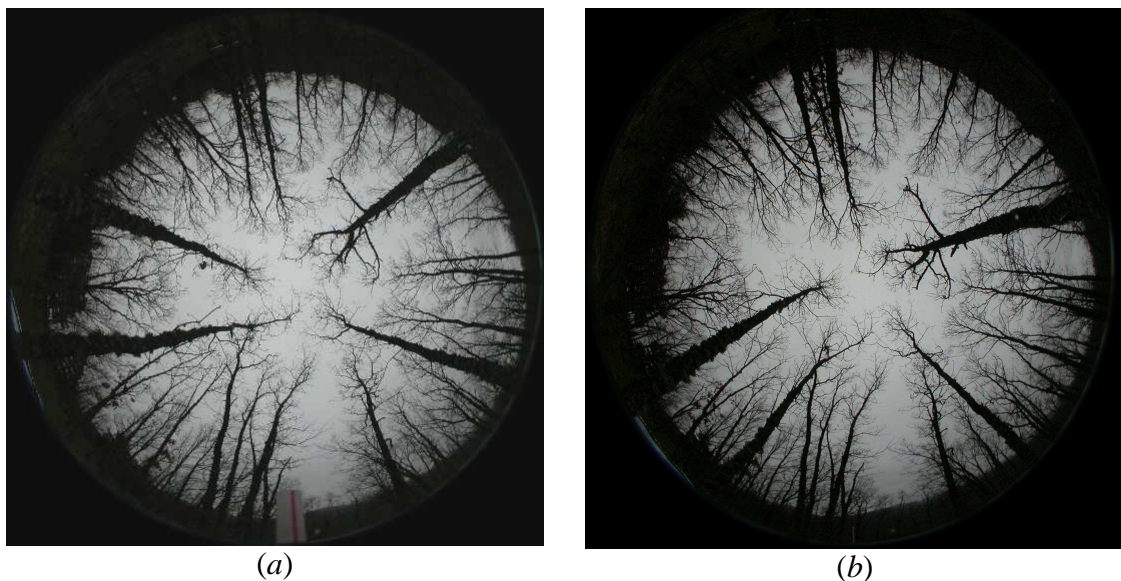


Figura 1.5 Imágenes originales omnidireccionales de un par estereoscópico captadas con baja iluminación  
(a) Imagen izquierda; (b) Imagen derecha.

Por otro lado, a pesar de la apariencia en tonalidades de gris de las imágenes representadas en la figura 1.5(a) y (b), lo cierto es que dichas imágenes están representadas en el modelo de color *RGB*. Este hecho se pone de manifiesto cuando sobre las mismas se realiza un proceso de realzado. En efecto, si sobre dichas imágenes realizamos una transformación desde el modelo *RGB*, que es como se encuentran originalmente, al modelo *HSI* (*Hue-matiz, Saturation-saturación, Intensity-intensidad*)



y sobre la componente de intensidad  $I$  se realiza un proceso de realzado mediante igualación uniforme de histograma (Pajares y Cruz, 2007a; Gonzalez y Woods, 2008), el resultado es el mostrado en las figuras 1.6(a) y (b) respectivamente. Como puede apreciarse, ahora las componentes espectrales se manifiestan explícitamente. A la vista de lo cual, podría pensarse en aprovechar el resultado de este realzado para aplicar un proceso de segmentación y correspondencia similar al propuesto para el primer tipo de imágenes, con vistas a diseñar un único proceso con validez general.

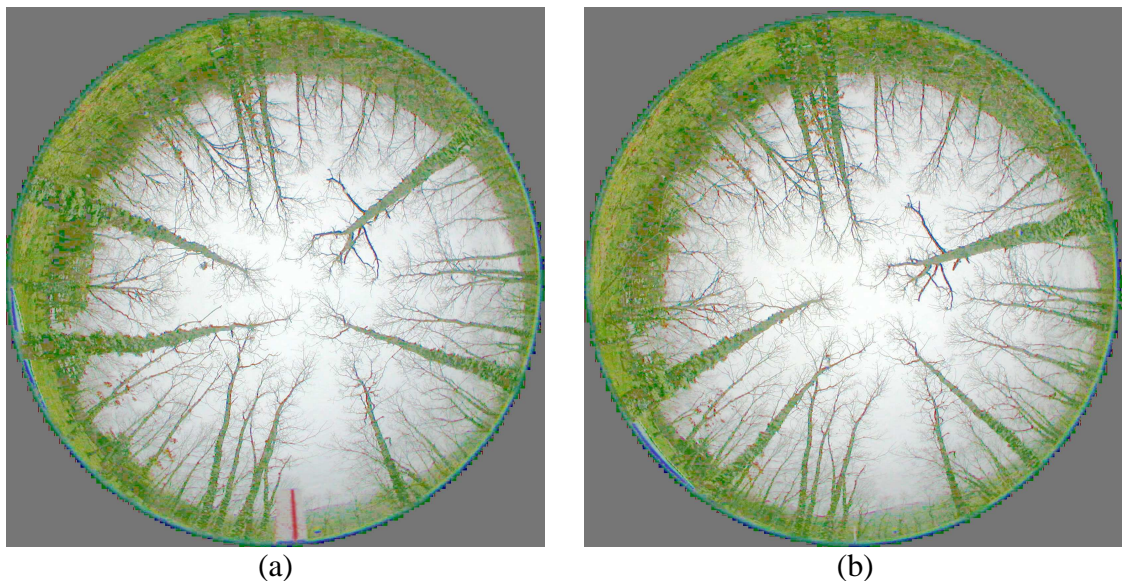


Figura 1.6 Imágenes omnidireccionales del par estéreo en la figura 1.5 realizadas por igualación de histograma uniforme (a) Imagen izquierda. (b) Imagen derecha.

Sin embargo, basándose en la observación de las imágenes, representadas en la figura 1.5, o equivalentemente en la figura 1.6, se pueden derivar los siguientes detalles útiles que aconsejan el diseño de una estrategia específica para este tipo de imágenes.

- 1) En la parte central de la imagen, hasta un determinado nivel, el cielo y los troncos se distinguen fácilmente por su contraste. En este tipo de entornos forestales y bajo las condiciones de iluminación descritas, la transmisión de la luz directa y difusa en todas las direcciones del hemisferio procedente del cielo permite obtener una imagen bien contrastada, en comparación con los sistemas convencionales (Wulder y Franklin, 2003). Lamentablemente esto no ocurre en la parte exterior de la circunferencia válida de la imagen debido a que tanto los troncos como la hierba en el suelo presentan firmas espectrales similares en el espacio de color  $RGB$ , siendo gris oscuro en la imagen en la figura 1.5 y verde en la imagen realzada en la figura 1.6. Por consiguiente, no se obtiene ninguna



ventaja apreciable por el hecho de utilizar las imágenes realzadas frente a las originales.

- 2) Los árboles están limpios de hojas en las ramas, lo que facilita su identificación, no requiriendo descriptores de textura con esta finalidad, como ocurría en las imágenes de los pinos con hoja perenne.
- 3) Los troncos presentan una cierta orientación hacia el centro; esto quiere decir que en la escena 3D son casi verticales. Sin embargo, algunos de ellos no son exactamente verticales e incluso pueden aparecer formas caprichosas como uno fácilmente puede observar en cualquier bosque de esta naturaleza. Esto debe tenerse en cuenta ya que nos impide aplicar exactamente las propiedades geométricas radiales durante su procesamiento. Hecho éste que se había puesto de manifiesto también en el caso del tipo de imágenes descrito en la sección previa.
- 4) Como ya se ha puesto de manifiesto previamente, debido a que las imágenes se capturan por la ubicación de la cámara en dos posiciones separadas por la línea base, 1m en nuestro caso, el mismo árbol de la escena 3D no se encuentra en la misma posición espacial en ambas imágenes. Se observa claramente, sin más que observar cualquiera de los árboles, un desplazamiento relativo, medido en grados de ángulo, entre los troncos de las imágenes que representan el mismo árbol 3D. Este desplazamiento es tanto mayor cuanto más cerca se sitúan los árboles del dispositivo estereoscópico.
- 5) Dependiendo de la posición de cada árbol con respecto a cada cámara, las imágenes de los troncos aparecen bajo diferentes tamaños, afectando al área del tronco proyectado en ambas imágenes. Un árbol próximo a la ubicación de la cámara cuando se toma la imagen izquierda aparece con un área en dicha imagen mayor que su área en la imagen derecha y viceversa.

De las observaciones realizadas previamente se deducen las siguientes pautas, que son determinantes para el diseño del procedimiento propuesto:

- 1) Gracias al alto contraste que aparece en la parte central, es posible distinguir fácilmente entre troncos y cielo, por lo que aprovechando esta circunstancia

resulta relativamente fácil proceder a la identificación de los troncos en esta zona, mediante técnicas de segmentación basadas en crecimiento de regiones (Pajares y Cruz, 2007a). Tras lo cual se obtiene una cierta intuición sobre el progreso que cada tronco va a seguir hasta su entroncamiento con el suelo. Esto se ve facilitado enormemente por el hecho de que los árboles carecen de hojas en sus ramas.

- 2) Además, como en el caso de los pinos, a pesar de presentar una orientación aparentemente radial, realmente esto no es cierto, presentando desviaciones en este sentido más agudizadas si cabe, que en el caso de los pinos por la naturaleza de este tipo de bosques, eso sin tener presente las formas caprichosas, que a veces se presentan. En consecuencia, tampoco son factibles las técnicas basadas en la orientación radial de los troncos.
- 3) Debido a las diferencias de tamaño entre las imágenes de un mismo tronco en el espacio 3D, así como a los relativamente altos valores de disparidad, una buena aproximación puede consistir en utilizar las regiones, que identifican los troncos, como características para la correspondencia. Esto se ve favorecido en este caso por el hecho de que las regiones ofrecen mayor robustez, que los píxeles individuales, por la sencilla razón de que representan agrupaciones compactas de píxeles con alguna propiedad común, tal como la intensidad.

Como conclusión al estudio anterior, un buen enfoque para el proceso estereoscópico puede orientarse a la extracción de regiones basándose en los puntos 1 y 2 previos. De estas regiones se pueden extraer una serie de propiedades, cuyo fundamento surge del punto 3; llegando, de este modo, a identificar dos procesos como en el caso anterior, esto es: *segmentación* y *correspondencia*. El primero identifica las regiones pertenecientes a los troncos, junto con sus propiedades y el segundo se encarga de emparejamiento de dichas regiones para determinar su disparidad.

En definitiva, tanto para el primer tipo de imágenes como para el segundo, el trabajo de investigación desarrollado en esta tesis se centra en el diseño de métodos y estrategias para llevar a cabo los procesos de segmentación y correspondencia que constituyen los pasos esenciales de todo proceso de visión estereoscópica, tal y como se

ha mencionado previamente y se describe en la literatura especializada, que se revisa en el capítulo dos.

## **1.3 Motivación y objetivos**

### **1.3.1 Motivación**

El trabajo de investigación desarrollado tiene su origen en el interés suscitado por el Centro de Investigación Forestal (CIFOR) del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) para automatizar los procesos de medición mediante el dispositivo con número de patente MU-200501738, mencionado previamente y descrito con detalle en el capítulo cinco.

De la necesidad suscitada por parte del CIFOR para explotar la potencialidad de dicho dispositivo de medida de cara a la automatización de los procesos para determinar medidas útiles en inventarios forestales, surgió un proyecto de colaboración entre el grupo ISCAR (2006) y el CIFOR titulado *Análisis estereoscópico y reconocimiento de texturas en imágenes forestales obtenidas mediante el dispositivo de medición MU-200501738* al amparo del artículo 83 de la LOU con Ref. INIA 445/2007. Sobre la base de este proyecto se mantiene una estrecha colaboración entre el grupo ISCAR y el CIFOR, con el propósito de perfeccionar y ampliar el campo de investigación en el mayor número de entornos forestales posible.

Por otro lado, el grupo ISCAR (2006) mantiene abierta una línea de investigación en tareas relativas a la visión estereoscópica, como consecuencia del desarrollo de varios proyectos donde dichas tareas se incluyen como fundamentales. Dado que el tipo de imágenes estudiadas en el presente trabajo de investigación, plasmado en esta memoria de tesis, son mucho más complejas y por tanto con un mayor grado de dificultad desde el punto de vista de su procesamiento, que las utilizadas en los citados proyectos, basadas en sistemas estereoscópicos convencionales, resulta claro que las propuestas planteadas aquí son fácilmente aplicables en los sistemas convencionales. Esta es una razón por la cual el grupo ISCAR al que pertenezco actualmente como miembro ha optado por mantener la investigación en el ámbito forestal objeto de esta memoria. Además, también participo como investigador en el proyecto del Plan Nacional de I+D que se detalla en el punto uno siguiente, sobre el que

está previsto adaptar alguna de las estrategias aquí planteadas. A continuación se detallan las líneas generales de tales proyectos, incidiendo en los objetivos relacionados con la visión estereoscópica:

1. Proyecto financiado por la CICYT titulado *Sistema de Vigilancia, Búsqueda y Rescate en el Mar mediante colaboración de Vehículos Autónomos Marinos y Aéreos* (DPI2009-14552-C02-01). En este proyecto se plantea como una de las tareas fundamentales la colaboración de vehículos tanto aéreos como terrestres o marinos. En todos los casos se hace necesario establecer el posicionamiento relativo entre ellos, siendo de especial relevancia el que se refiere al posicionamiento de barcos para determinar sus coordenadas geográficas a partir de un punto de referencia que establece las coordenadas base. Este tipo de localización se está llevando a cabo mediante técnicas de visión estereoscópica con el fin de corregir desviaciones introducidas por otros sistemas tales como el GPS. El objetivo final del proyecto consiste en la colaboración de los diversos vehículos para su actuación e intervención en situaciones complicadas tales como intervención en vertidos de petróleo en el mar, incendios forestales u otras eventualidades.
2. Durante el trienio 2006-2008 se ha colaborado en el subproyecto CICYT titulado *Extracción de Conocimiento Espacio-Temporal y Visión Artificial para la Automatización de Tratamientos Localizados* (AGL2005-06180-C03-03) con el Instituto de Automática Industrial del CSIC. Este proyecto tiene su continuidad en un nuevo proyecto del Plan Nacional de I+D 2008 para el trienio 2009/2011 concedido con el número de referencia AGL2008-04670-C03-02/AGR, con el título *PERcepción visual, toma de deCisión y Actuación en sistemas artificiales complejos: control selectivo de malas hierbas en cultivos agrícolas (PERICIA)*. En ambos proyectos se incorporan técnicas de visión estereoscópica para la localización espacial de rodales de malas hierbas en campos de cultivo de cereal y maíz con el fin de proceder a su pulverización selectiva. El objetivo final del proyecto consiste en minimizar el efecto contaminante medioambiental por el tratamiento con herbicidas de los campos de cultivo, así como disminuir los costes de producción.

3. Con la empresa *TCP Sistemas e Ingeniería* se colabora en dos proyectos titulados respectivamente *AUTOROVER: estudio de autonomía basada en imágenes para “rover” de exploración planetaria* y *Visión estereoscópica para Auto-rover: estudio de autonomía basada en imágenes*. El primero procede de la participación de la citada empresa en la convocatoria pública según la orden 2259/2007 (B.O.C.M. 272 de 15/11/2007) para el fomento de la innovación en el sector aeroespacial de la Comunidad de Madrid con financiación del Fondo Europeo Regional (Referencia 04-AEC0800-000035/2008). El segundo corresponde a una ampliación del anterior dentro del programa Nacional de Proyectos de Investigación en el Sector Aeroespacial según la orden PRE/998/2008 (BOE 11/04/2008) del Ministerio de la Presidencia (Referencia SAE-20081093). En ambos proyectos el objetivo consiste en determinar las distancias a las que se encuentran los objetos mediante técnicas de visión estereoscópica de cara a la navegación autónoma de los vehículos de exploración planetaria.

Aunque bien es cierto, que los proyectos descritos en los puntos 1, 2 y 3 previos, están basados en sistemas estereoscópicos convencionales cuyas imágenes se obtienen mediante proyección de perspectiva, no es menos cierto que las técnicas investigadas en esta tesis, particularmente las de correspondencia estereoscópica a nivel de píxel, son aplicables a los sistemas convencionales, por lo que se espera una fácil reutilización de los métodos y estrategias desarrollados en este trabajo para cámaras basadas en lentes de ojo de pez, para su aplicación a los sistemas que utilizan cámaras convencionales.

Además, desde el punto de vista tecnológico, la motivación viene suscitada por el hecho de automatizar una tarea que en la actualidad se realiza de manera manual con la mejora que eso supondría en cuanto a ahorro de costes y disminución del impacto medioambiental, y naturalmente por el desarrollo de procedimientos como respuesta a la creciente demanda tecnológica en Visión por Computador.

### 1.3.2 Objetivos

A la vista de las consideraciones expresadas en los apartados precedentes, se plantean los siguientes objetivos de investigación relativas a los procesos de segmentación y correspondencia previstos:

1. Estudio bibliográfico de la literatura existente a nivel de segmentación de imágenes, abordando el tema desde la clasificación de texturas hasta la identificación y descomposición de regiones.
2. Estudio bibliográfico de métodos de correspondencia en imágenes en sistemas de visión convencionales basados en sistemas ópticos con proyección de perspectiva y sistemas de visión omnidireccionales equipados con lentes de ojo de pez.
3. Diseñar estrategias de segmentación de las imágenes a nivel de píxel y a nivel de características relevantes, tales como regiones.
4. Diseñar métodos de correspondencia estereoscópica que trabajen a nivel de píxeles y a nivel de regiones.
5. Analizar los resultados extrayendo las pautas necesarias para la identificación y correspondencia de los troncos.
6. Identificar las líneas de investigación futuras, tanto desde el punto de vista de la mejora de las estrategias propuestas como desde el diseño de nuevos métodos para su aplicación al mayor número posible de entornos forestales.
7. Divulgar y difundir los resultados de la investigación mediante publicaciones científicas u otros mecanismos.

## 1.4 Metodología

En el desarrollo del trabajo se han seguido las directrices marcadas por el método científico. Inicialmente, se ha realizado una búsqueda de información bibliográfica en todas las fuentes disponibles, utilizando principalmente para ello el acceso a través de la red del Consorcio Madroño al que están suscritas las Universidades de la Comunidad de Madrid. Se han estudiado los antecedentes metodológicos y se ha analizado y evaluado el estado del arte en este área. Seguidamente, se han dado los pasos que se indican a continuación, a la vez que se han revisado de manera continua las fuentes bibliográficas hasta el final de la investigación e incluso en el momento de la redacción de este trabajo. Las etapas seguidas han sido:

1. Definición del problema de investigación en función tanto de las necesidades identificadas como de la motivación establecida previamente.
2. Planteamiento de los objetivos de la investigación.
3. Formulación de planteamientos verificables.
4. Obtención de datos experimentales.
5. Clasificación, análisis e interpretación de los datos.
6. Conclusiones generales.
7. Descripción de líneas futuras de investigación.

Para la realización de los experimentos se han implementado los métodos basados en las técnicas propuestas. Este proceso abarca desde la adquisición de datos para la creación de una base suficientemente amplia con la que experimentar, hasta la obtención de resultados, pasando por su visualización. Las imágenes utilizadas en los experimentos son cortesía del CIFOR.

## 1.5 Aportaciones de la investigación

La contribución principal de este trabajo consiste en la combinación de una propuesta de segmentación y un proceso de correspondencia en entornos forestales para dos tipos de imágenes omnidireccionales obtenidas con lentes de ojo de pez y con características diferentes: *a)* alta variabilidad de la intensidad en los troncos en función de la incidencia del sol; *b)* imágenes con pobre iluminación y alto contraste entre el cielo y los troncos en la parte central. Ambos procesos se han diseñado bajo la perspectiva de su automatización.

Para el primer tipo de imágenes analizadas la contribución se ha materializado en los dos procesos diseñados. De este modo, en cuanto a la segmentación se propone la combinación de dos clasificadores con el fin de identificar y excluir las texturas mencionadas, estos clasificadores son: *Agrupamiento Borroso (FC)* y el *estimador paramétrico Bayesiano (PB)*. Para el proceso de correspondencia se proponen varias estrategias enmarcadas en la teoría de toma de decisiones para obtener en una primera fase un mapa de disparidad inicial. Dichas estrategias se refieren exactamente a la

adaptación de los siguientes paradigmas: *Integral Fuzzy de Choquet (CFI)*, la *Integral Fuzzy de Sugeno (SFI)*, la *Teoría Dempster-Shafer (DES)* y la *Toma de Decisiones Multicriterio bajo el paradigma de la lógica Fuzzy (MCDMF)*. Este mapa de disparidad inicial, es posteriormente mejorado, en una segunda fase, mediante la aplicación de dos estrategias iterativas, a saber: los *Mapas Cognitivos Fuzzy (FCMs)* y la *Red Neuronal de Hopfield (HNN)*; ambas son dos estrategias muy utilizadas en tareas y procesos relacionados con la Visión por Computador. Los métodos propuestos se han comparado favorablemente en el ámbito de los desarrollos realizados.

En relación al segundo tipo de imágenes analizadas, la principal contribución consiste en el diseño de un proceso de segmentación específico orientado hacia la extracción de regiones como paso previo a la descripción de las mismas mediante un conjunto de propiedades, que permitirán la correspondencia entre regiones de ambas imágenes del par estereoscópico. Ambos, segmentación y correspondencia, al igual que en el caso previo, definen la estrategia global en este tipo de imágenes. En este caso, el comportamiento de la estrategia propuesta se compara favorablemente frente al criterio del experto humano, que procesa las imágenes basado en su pericia, si bien manualmente.

Los resultados obtenidos hasta el momento, que han permitido su divulgación a nivel internacional, son los que se recogen a continuación:

1. En Herrera y col. (2009a) se plantea una propuesta para establecer correspondencias entre píxeles mediante el paradigma *CFI*, que es una técnica aplicada con éxito en tareas relacionadas con la combinación de clasificadores. De esta forma se obtiene un mapa de disparidad inicial, que contiene las disparidades de cada píxel de la imagen a nivel individual. Este trabajo se encuentra publicado en la serie *Advances in Soft Computing* de la editorial Springer-Verlag.
2. En Herrera y col. (2009b) se proponen dos estrategias combinadas: *SFI* y *DES*. El diseño de la combinación aplicado a nuestro problema de correspondencia estereoscópica basada en píxeles, resulta ser la principal contribución de este trabajo, publicado en la serie *Lecture Notes in Computer Science* de la editorial



Springer-Verlag. Mediante esta estrategia combinada se obtiene un mapa de disparidad inicial como en el trabajo previo.

3. En Herrera y col. (2009c) se diseña un proceso de correspondencia estereoscópica también basado en píxeles mediante el paradigma *MCDMF*, obteniendo igualmente un mapa de disparidad como resultado de su aplicación. Este trabajo se encuentra publicado en la serie *Lecture Notes in Computer Science* de la editorial Springer-Verlag.
4. En Pajares y col. (2009) y Guijarro y col. (2008, 2009) se proponen sendas estrategias relativas a la identificación y clasificación de texturas en entornos naturales. Las imágenes sobre las que se han aplicado son imágenes procedentes de sensores aerotransportados. Las imágenes forestales utilizadas en el presente trabajo y aquellas tienen en común que son imágenes de entornos naturales de exterior y por tanto no estructurados. Los resultados de investigación realizados sobre aquéllas han servido de base para la propuesta de segmentación en las imágenes de pinos utilizadas en este trabajo. En concreto, la propuesta de combinación de clasificadores formulada en dicho proceso de segmentación proviene de esas investigaciones. Es más, al llevarse ambas líneas de investigación en paralelo ha existido un aprovechamiento mutuo entre ambas. El primer trabajo se encuentra publicado en la revista *IET Computer Vision* con índice de impacto de 0.667 en 2008, encontrándose posicionada en el segundo tercio de su categoría. El segundo trabajo está publicado en la serie *Lecture Notes in Computer Science* bajo la subserie *Lecture Notes in Artificial Intelligence* de la editorial Springer-Verlag. El tercer trabajo se encuentra publicado en la revista *Sensors* con índice de impacto de 1.870 en 2008, encontrándose posicionada en el primer tercio de su categoría.
5. En Herrera y col. (2009d) se describe un novedoso proceso en correspondencia estereoscópica basado en regiones. Los pasos clave son la extracción de características y su correspondencia. En una primera etapa existe un proceso de segmentación que extrae las regiones asociadas con los troncos, éstas constituyen las características bajo correspondencia, de las regiones se obtienen una serie de propiedades con las que se establecen las correspondencias entre características. La combinación de los procesos de segmentación y

correspondencia para esta clase específica de sensores son la contribución principal de este trabajo que se encuentra publicado en la revista *Sensors* con índice de impacto de 1.870 en 2008, encontrándose posicionada en el primer tercio de su categoría.

En resumen, los trabajos de las referencias citados en los puntos 1, 2 y 3 constituyen la base de los procesos de correspondencia para el análisis de las imágenes procedentes de los pinos, mientras que las del punto 4, como ya se ha mencionado, constituyen el punto de partida y referencia del proceso de segmentación en dicho tipo de imágenes. Finalmente, la referencia citada en el punto 5, resulta ser el fundamento y el núcleo de las propuestas de segmentación y correspondencia formuladas en la presente memoria de tesis para el segundo tipo de entorno forestal de robles analizado.

## 1.6 Organización de la memoria de tesis

La memoria de tesis se presenta estructurada en capítulos. Estos capítulos se organizan siguiendo el orden natural de la investigación, cuya distribución es la que se ofrece a continuación:

Primer capítulo. *Introducción*, donde se estudian y plantean los objetivos de la tesis, así como la motivación e introducción de las propuestas y planteamientos realizados, tal y como se ha expuesto en las secciones precedentes.

Segundo capítulo. *Estado del arte*, donde se revisan diferentes estrategias existentes en la literatura, que proporcionan los argumentos y la base de la investigación.

Tercer capítulo. *Segmentación*, en el que se distinguen dos partes. En la primera, orientada al enfoque basado en píxeles, se describen los procedimientos *FC* y *PB* aplicados a la segmentación de las imágenes para la identificación de las texturas a ser excluidas durante el proceso de correspondencia en las imágenes de pinares. En la segunda parte, orientada al enfoque basado en regiones, se describe el procedimiento específico aplicado para la segmentación de la imagen orientada a la identificación de las regiones que describen a los troncos en las imágenes de robles.

Cuarto capítulo. *Correspondencia estereoscópica*. A nivel de píxeles podemos distinguir dos partes, la primera describe el diseño del proceso de correspondencia basado en los paradigmas *CFI*, *SFI*, *DES* y *MCDMF* con los que se obtiene un mapa de disparidad inicial. La segunda parte describe los paradigmas *FCM* y *HNN*, cuya finalidad es la mejora de ese mapa de disparidad inicial. A nivel de regiones se describe el diseño del proceso de correspondencia basado en dichas características.

Quinto capítulo. *Análisis de resultados*, donde se presentan los resultados obtenidos. En el enfoque basado en píxeles se comparan los resultados obtenidos por las diferentes estrategias, así como el efecto y la incidencia del proceso de segmentación en el procedimiento total. En el enfoque basado en regiones, la comparación se establece frente al criterio del experto humano, en este caso considerando ambos procesos de segmentación y correspondencia conjuntamente.

Sexto capítulo. *Conclusiones y Trabajo futuro*, donde se exponen las líneas generales seguidas, realizando un balance general del trabajo. Se plantean las líneas de investigación futuras así como posibles mejoras o extensiones de los métodos propuestos.

Apéndice A. *Sistemas de Visión Estereoscópica*, donde se realiza un amplio recorrido sobre diversos métodos y sistemas cuyo fundamento es precisamente la visión estereoscópica, si bien desde un punto de vista relacionada con los sistemas convencionales, como lo es el sistema visual humano.

Apéndice B. *Segmentación y correspondencia: nuevos retos*. En aras de la extensión de las estrategias propuestas a otros entornos forestales, se realiza un avance sobre las futuras líneas de investigación que sería conveniente seguir para la aplicación de sendas estrategias de segmentación y correspondencia a un bosque de robles en condiciones de iluminación similares a las de los pinos. Entre otras cuestiones se exponen las razones fundadas que justifican la propuesta de un procedimiento de segmentación previo al de correspondencia.

## Estado del arte

### 2.1 Visión general

En el capítulo uno se ha establecido como objetivo principal la automatización del proceso para la obtención de las distancias a puntos significativos de la escena 3D desde el dispositivo de medición MU-200501738 con el fin de poder estimar una serie de parámetros tendentes al análisis y la realización de inventarios forestales.

Como también se ha mencionado en el capítulo previo, dicho dispositivo es exactamente un sistema estereoscópico basado en imágenes, por tanto, el proceso de automatización pasa por el procesamiento de dichas imágenes bajo el ámbito de cualquier proceso basado en visión estereoscópica pasiva. Aunque las imágenes poseen las características propias de los sistemas ópticos basados en lentes de ojo de pez, el proceso estereoscópico consta de los pasos propios de cualquier sistema estereoscópico convencional basado en proyección de perspectiva, sistemas éstos que han sido por tradición los estudiados ampliamente en la literatura, en contraposición a los basados en lentes de ojo de pez sobre los que la base documental resulta ser más bien escasa, hasta donde hemos podido investigar, razón por la cual el análisis bibliográfico realizado en este capítulo se basa en la información relativa a los sistemas estereoscópicos convencionales.

Por tanto, independientemente del sistema utilizado, convencional o basado en ojos de pez, siguiendo la terminología establecida en Barnard y Fishler (1982) o Cochran y Medioni (1992), el proceso de la visión estereoscópica consta de los siguientes pasos:

- a)* adquisición de imágenes,
- b)* geometría del sistema,
- c)* extracción de características,
- d)* correspondencia de características.
- e)* obtención de la distancia.

La adquisición de imágenes es el proceso por el cual se obtienen las dos imágenes del par estereoscópico, en nuestro caso las imágenes con la proyección propia de los sistemas basados en lentes de ojo de pez, tal y como se mostraron en el capítulo uno, figuras 1.3 y 1.5.

La geometría del sistema determina las relaciones físico-geométricas del dispositivo, que definen ciertos parámetros que afectan al proceso global de la visión estereoscópica. La geometría del dispositivo que genera las imágenes de esta investigación, se detalla en el capítulo cuatro.

La extracción de características se refiere al proceso por el cual se identifican en las imágenes los elementos propios de las mismas, que se emparejarán entre ambas imágenes y que representan las proyecciones de los elementos reales en la escena 3D de suerte que un elemento, por ejemplo un árbol en la escena, genera sendas proyecciones en las dos imágenes del par estereoscópico. Dichas características pueden ser de diversos tipos, desde los píxeles que forman la imagen en general, pasando por píxeles específicos como los que delimitan puntos de borde hasta agrupaciones de píxeles formando regiones según alguna propiedad común de los mismos. Desde el punto de vista de la investigación desarrollada en este trabajo y por las razones expuestas en el capítulo uno, las características utilizadas serán píxeles para las imágenes procedentes de los pinares con alta variabilidad de iluminación, y regiones para las imágenes del bosque de robles con escasa iluminación. En ambos casos es necesario extraer de cada característica una serie de propiedades o atributos, de suerte que por similitud entre ellos puedan hacerse corresponder características de una imagen con las características de la

otra imagen. Por otro lado y de forma especial para el tipo de imágenes procedentes de los pinares, donde las preferencias estriban en identificar los píxeles pertenecientes a los troncos de los árboles, también se incluyen como características aquellos píxeles que no forman parte de los troncos sino de aquellas otras texturas cuyo objetivo, dentro del proceso, es su exclusión, mediante las técnicas de clasificación previstas, del proceso de emparejamiento y favorecer así el emparejamiento de los troncos. Al proceso completo de extracción de características, extracción de propiedades de las características y clasificación de texturas para su exclusión se le identifica, desde el punto de vista de esta investigación, como proceso de *segmentación*, tal y como se había adelantado previamente en el capítulo uno.

La correspondencia de características es el proceso por el cual dada una característica en la escena 3D se llega a determinar cuál es su proyección en sendas imágenes del par estereoscópico. Es decir, se trata de emparejar características homólogas de una imagen con las de la otra imagen del par estereoscópico, que representan la misma entidad física en la escena 3D. Para lograr los emparejamientos, se aplican una serie de restricciones conocidas en la literatura especializada como *restricciones estereoscópicas*. Este proceso se conoce como *correspondencia* propiamente dicho. Una vez establecida la correspondencia entre dos características, resulta fácil determinar las ubicaciones en las imágenes de ambas características y de ahí determinar las separaciones relativas entre ellas, esto es, la disparidad, que en el caso de las imágenes analizadas se traduce en diferencias entre los dos ángulos que determinan las posiciones de las características en coordenadas polares, como se detalla en los capítulos tres y cuatro.

Finalmente, una vez establecidas las correspondencias entre características homólogas y por tanto sus disparidades, resulta relativamente fácil determinar las distancias a las que se encuentran en la escena 3D las entidades representadas por dichas características sin más que recurrir a ciertos parámetros conocidos y proporcionados por la geometría del sistema estereoscópico, tal y como se describe en el capítulo cuatro.

De lo expuesto anteriormente, son los procesos de segmentación y correspondencia los que verdaderamente constituyen la clave del proceso global de correspondencia estereoscópica en general y en particular para el tipo de imágenes utilizadas en este trabajo de investigación. Por este motivo, la investigación que se

presenta se centra exclusivamente en proporcionar una serie de estrategias y procedimientos encaminados a dar solución tanto al problema de la segmentación como al de la correspondencia para los dos tipos de imágenes descritas en el capítulo uno. En la figura 2.1 se presenta un diagrama esquemático del proceso global de visión estereoscópica, donde como puede observarse bajo el epígrafe de segmentación de imágenes, se agrupan los procesos de extracción de características y de los atributos de dichas características, cuya salida constituye la entrada al proceso de correspondencia.

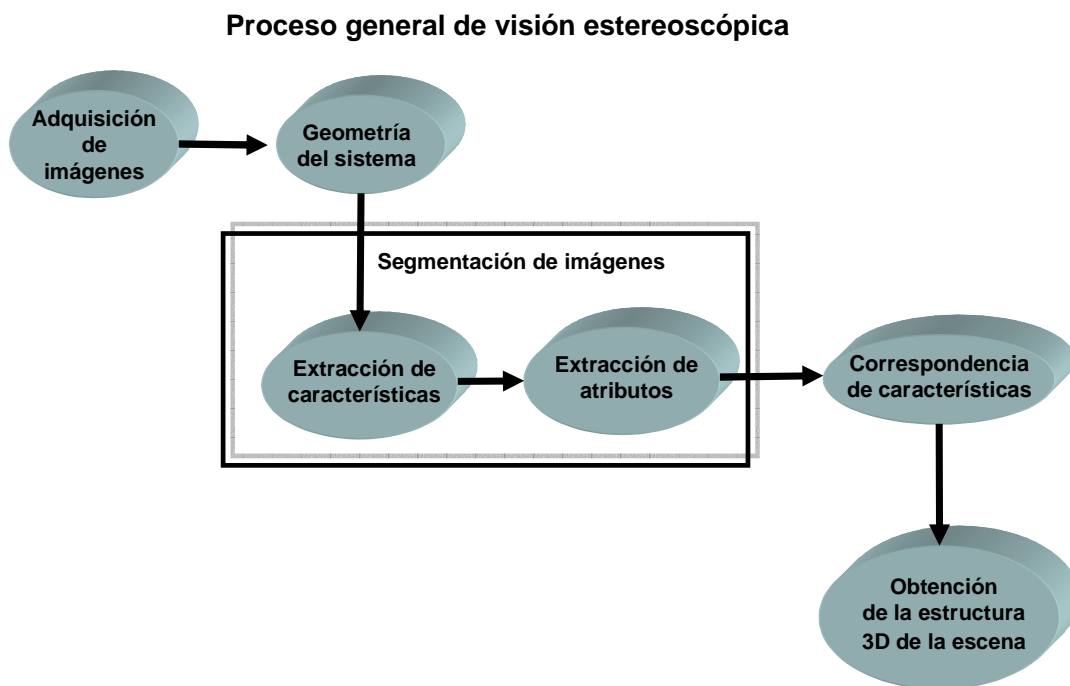


Figura 2.1 Esquema del proceso de visión estereoscópica.

En resumen, el trabajo de investigación que se presenta aborda de forma directa y específica los procesos de segmentación y correspondencia.

En el proceso de segmentación se encuadran aquellas técnicas encaminadas a la identificación de características tanto a nivel de píxel como a nivel de regiones. En ambos casos, la segmentación constituye un paso clave y esencial. En efecto, en el enfoque basado en regiones ésta es la parte que entraña la mayor dificultad dentro del proceso, como se expone en el capítulo tres; pero aún si cabe, en el enfoque basado en píxeles resulta ser tanto o más importante. En este caso su objetivo se centra en la identificación de los píxeles pertenecientes a una serie de texturas para su exclusión, que facilitan sin duda el proceso de correspondencia relativo a los píxeles de los troncos. Este hecho se pone de manifiesto en los resultados mostrados en el capítulo cinco. Aún

es más, en determinados tipos de imágenes el éxito o fracaso de la correspondencia está ligado directamente a la correcta exclusión o no de las texturas carentes de interés, tal y como queda patente en el análisis realizado en el apéndice B sobre las imágenes procedentes de bosques de robles con hojas y en condiciones de iluminación similares a las de los pinos. Como se refleja en el mencionado apéndice, el problema de la segmentación en este tipo de imágenes permanece abierto a la investigación, por este motivo resulta absolutamente imprescindible realizar un estudio bibliográfico en profundidad sobre métodos de identificación y clasificación de texturas en entornos naturales, ya que hasta donde se ha podido investigar no existen técnicas específicas que den solución a dicha problemática. También se debe justificar dicho estudio bibliográfico sobre la base de la búsqueda de mejoras en las propuestas formuladas en este trabajo.

En la sección 2.2.1 se revisan distintos métodos genéricos encaminados al reconocimiento de texturas. La sección 2.2.2 realiza un recorrido por distintos métodos orientados a la clasificación de texturas en imágenes naturales, tanto a nivel de píxel como de regiones. Específicamente, se incluyen tanto las propiedades espectrales (color), como espaciales (distribuciones de texturas) o ambas bajo un enfoque combinado. Posteriormente, en la sección 2.2.3 se analizan en concreto métodos cuyo fundamento son las relaciones espaciales de vecindad existentes entre un píxel y los que le rodean. Aquí se contemplan las agrupaciones irregulares espaciales de píxeles, que son realmente el tipo de agrupaciones de píxeles que aparecen en las imágenes naturales objeto de este trabajo de investigación.

El proceso de correspondencia engloba todos los aspectos encaminados a realizar el emparejamiento de características. Desde esta perspectiva, en primer lugar, en la sección 2.3.1 se realiza una revisión sobre las restricciones comúnmente aplicadas en visión estereoscópica, dando detalles sobre cuáles de ellas son las más apropiadas para el tipo de imágenes utilizadas en nuestra investigación. En segundo lugar, en la sección 2.3.2 se estudian las técnicas de correspondencia cuyo fundamento son los píxeles de forma individual o las agrupaciones de píxeles, en este último caso bien sea formando agrupaciones con cierto significado como podrían ser los troncos de los árboles en las imágenes o agrupaciones por similitud de píxeles vecinos. En ambos casos se recurre a la literatura general relativa a los sistemas estereoscópicos



convencionales con proyección de perspectiva, dado que para los sistemas con óptica basada en lentes de ojo de pez y hasta donde hemos podido indagar, la literatura resulta más bien escasa.

Finalmente, como consecuencia de la revisión previa, en la sección 2.4 se incluye un resumen sobre aquellas técnicas más relevantes desde el punto de vista del trabajo que se presenta, constituyendo por tanto, el origen de la investigación desarrollada y que se detalla en capítulos sucesivos.

## **2.2 Revisión de técnicas en clasificación de texturas**

### **2.2.1 Metodologías de clasificación de texturas en imágenes naturales**

Durante las últimas tres décadas e incluso cuatro, se han propuesto en la literatura diversos métodos encaminados a la segmentación de las imágenes según sus texturas, ver por ejemplo Xia y col. (2007) y referencias asociadas. En Wilkinson (2005) se realiza una exhaustiva revisión de métodos de clasificación durante los quince años previos a la publicación de dicho trabajo aplicados a las imágenes de satélite.

La mayoría de ellos pueden considerarse como métodos con dos procesos sucesivos asociados: *extracción de características* y *segmentación basada en las características*. El término sucesivos hace referencia al hecho de que se realizan en ese orden durante la clasificación.

La extracción de características tiende a encontrar una medida apropiada para caracterizar la homogeneidad de cada región dentro de la imagen, estableciendo una asociación entre la región y una determinada textura (Valdovinos y col. 2005, Puig y García, 2006 o Hanmandlu y col. 2004). Algunas de tales características se analizan posteriormente.

La segmentación basada en las características consiste en el proceso de asignar a cada píxel de la imagen una etiqueta que lo identifique como perteneciente a una determinada clase. Además, se han diseñado y utilizado una amplia gama de algoritmos tanto de naturaleza supervisada como no supervisada. A continuación se realiza un esbozo de los más relevantes, si bien algunos de ellos serán estudiados con mayor

profundidad en el capítulo tres, dado que constituyen la base de la propuesta de extracción de texturas a nivel de píxel realizada en este trabajo.

### 1. *Técnicas borrosas o fuzzy*

Asumen el hecho de que los colores tienden a agruparse en la imagen formando clases que determinan la distinta naturaleza de las texturas. Dada la naturaleza *fuzzy*, en cada una de las clases, se admiten píxeles con cierta ambigüedad, en el sentido de que un mismo píxel podría pertenecer tanto a una textura como a otra según sus características. Este hecho se manifiesta más claramente en las zonas fronterizas de las texturas, donde se producen transiciones graduales en los valores de los píxeles de una zona a otra de la frontera. A la categoría *fuzzy* pertenece uno de los métodos más clásicos en clasificación como es el conocido Agrupamiento borroso o en inglés *Fuzzy Clustering*. Para la fase de entrenamiento utiliza un método de optimización, que permite conseguir agrupaciones compactas y resolver muchas de tales ambigüedades. En esta línea se encuentran los trabajos de Pajares y Cruz (2002), Pajares y col. (2002) o Lim y Lee (1990) entre otros.

Mohaddamzadeh y Bourbakis (1997) definen las regiones a segmentar como agrupaciones de píxeles que poseen similar color, mientras los cambios de región se manifiestan gradualmente como ya se ha indicado previamente. Para plasmar esta idea definen una función de naturaleza *fuzzy* en el espacio de color *RGB* (Pajares y Cruz, 2007a).

Cheng y col. (2001) utilizan el concepto de *homogram* para identificar regiones homogéneas en una imagen de color; a su vez definen el concepto de *homogeneidad fuzzy* para extraer regiones de similares características mediante la obtención de una serie de umbrales. Tras la segmentación vuelven a aplicar un criterio de similitud para fusionar regiones, se trata de un proceso similar al método de segmentación de regiones según la técnica de árboles cuaternarios (Pajares y Cruz, 2007a).

En el ámbito de las imágenes de satélite los siguientes trabajos utilizan métodos de clasificación basados en la lógica *fuzzy*: Seong y Usery (2001), Zhang y Foody (1998).

## 2. *Redes neuronales*

Por el carácter no lineal, que permite una mayor separabilidad en el espacio de características y su alta capacidad para desarrollar procesamiento en paralelo, las redes neuronales se han utilizado también en la clasificación de texturas en imágenes en general y en imágenes de satélite en particular (Mas y Flores, 2008).

Dentro de las redes neuronales se encuentra la red neuronal de Hopfield (Haykin, 1994), compuesta por una serie de neuronas o nodos conectados entre sí. Este tipo de red realiza un proceso de optimización basándose en la minimización de una función de energía hasta que la propia red adquiere un grado de estabilidad aceptable. En Huang (1999) se detectan los picos más significativos del histograma y se diseñan tres redes de Hopfield, una para cada banda de color, donde los nodos de cada red son los píxeles de la correspondiente banda espectral y las conexiones establecen similitudes entre ellos. En Campadelli y col. (1997) se diseña un algoritmo con  $M \times N \times S$  neuronas siendo  $M$  y  $N$  las dimensiones de la imagen y  $S$  el número de clases que se obtienen mediante el análisis previo del histograma. En este tipo de redes se determinan las conexiones entre los nodos mediante el cómputo de funciones de consistencia entre los píxeles, basándose en las etiquetas de los vecinos. Para ello, se requiere la realización de un proceso de clasificación y etiquetado previo.

La red retro-propagación ha sido utilizada en clasificación de imágenes naturales, incluyendo las imágenes de satélite, y cabe mencionar entre otros los trabajos de Pajares y col. (2001), Atkinson y Tatnall (1997) o Paola y Schowengerdt (1995), donde las componentes espectrales correspondientes a los tres colores básicos de cada píxel constituyen las entradas a la red, mientras que como salidas se establecen las clases a las que pertenecen dichos píxeles según un mecanismo de aprendizaje supervisado.

Los mapas Auto-Organizativos (*Self Organizing feature Maps, SOM*) han sido otra de las herramientas utilizadas en clasificación. Este es el caso de los trabajos desarrollados en Jiji y Ganesan (2008) o Lo y Pei (1999) entre otros. En ellos los centros de las clases se estiman según la correspondiente ley de aprendizaje.

### 3. *Cuantización vectorial (Learning Vector Quantization)*

Se trata de un procedimiento que puede ser de naturaleza supervisada o no supervisada. En su versión supervisada se determina de antemano el número de clases existentes en la imagen, mientras que en la no supervisada es a través del propio método como se establece el número de clases en función de un umbral fijado previamente. En cualquiera de las dos versiones, se trata de un proceso de aprendizaje, que obtiene los centros de las clases. Algunos trabajos basados en estos métodos, que además utilizan información del color, son los propuestos en Giacinto y col. (2000), Chan y col. (2003) o del Frate y col. (2007). Los trabajos de Lillo y col. (2007), Aiyer y col. (2005) o Pyun y col., (2002) realizan una cuantización vectorial utilizando propiedades de texturas obtenidas mediante la aplicación de la transformada de Fourier (Pajares y Cruz, 2007a; Gonzalez y Woods, 2008).

### 4. *Vecinos más próximos*

Este método se conoce en la literatura inglesa como *k-Nearest Neighbours* (*k*-NN). En ellos, dado un conjunto de clases, los nuevos píxeles a clasificar con sus características, se asignan a aquella clase que posea un número de *k* vecinos más próximos. Dicha proximidad se determina mediante alguna medida de similitud, por ejemplo una distancia. En Haapanen y col. (2004) se utiliza esta técnica para la segmentación de zonas boscosas en imágenes aéreas. En Muñoz-Marí y col. (2007) se utiliza este clasificador como referencia para el análisis comparativo de su método. En Singh y col. (2001) se aplica este método para clasificar nueve tipos diferentes de texturas.

### 5. *Estimas de máxima verosimilitud*

Consistentes en la estimación de una función de densidad de probabilidad, que constituye el objeto del aprendizaje. Tomando como referencia la función de densidad de probabilidad estimada, es posible determinar la probabilidad de pertenencia de una muestra dada a las diferentes clases. En esta línea se encuentran los trabajos de Frizelle y Moody (2001), Newsam y col. (2004), Xie y Tsui (2004), Trauwaert y col. (1991) o Benediktsson y col. (1990). Muñoz-Marí y col. (2007) utilizan también este clasificador para comparar los resultados obtenidos con su método.

## 6. *Algoritmos genéticos*

Utilizados como métodos de optimización que aprovechan la información espacial de la vecindad para definir las funciones objetivo, tal es el caso de Tso y Mather (1999).

## 7. *Máquinas de vectores soporte*

Estas técnicas están especialmente diseñadas para aplicaciones bi-clase, si bien se han propuesto ampliaciones para abordar el problema multi-clase. Tal es el caso de los trabajos propuestos en Camps-Valls y Bruzzone (2005), Camps-Valls y col. (2004) o Melgani y Bruzzone (2004). En Ferreiro-Armán y col. (2007) se realiza un estudio sobre el comportamiento en imágenes de satélite de las máquinas de vectores soporte, además del Agrupamiento borroso, los Mapas Auto-organizativos y la Red neuronal retro-propagación.

## 8. *Árboles de decisión*

Se presentan como una alternativa a los métodos clásicos de clasificación donde lo que se infieren son reglas de decisión, que permiten realizar las clasificaciones en base al cumplimiento de las premisas aprendidas y reflejadas en las reglas (Tottrup, 2007; Kumar y Majumder, 2001; Hansen y col. 1996). En Debeir y col. (2002) se utiliza el esquema de árbol conocido como C4.5 para la clasificación de imágenes aéreas.

Un aspecto importante a destacar respecto de los métodos mencionados en los puntos uno a ocho, es el hecho de que se centran exclusivamente en la clasificación de los píxeles sin tener en cuenta las características espaciales de las imágenes. Así, para cada píxel de la imagen se extraen las correspondientes propiedades o características, procediendo seguidamente a su clasificación sin tener en cuenta su posición o relación con otros píxeles vecinos en la imagen y por extensión en la escena 3D. Con el fin de incorporar la información espacial, se han realizado diferentes esfuerzos para considerar esta importante propiedad de las imágenes. En efecto, en cualquier imagen, y en particular en las analizadas en este trabajo, cada píxel se ubica en una determinada localización espacial dentro de la imagen. Después de extraer las propiedades de cada píxel, la imagen puede expresarse como agrupaciones de ciertas propiedades, de forma que cada agrupación define una región. Cada región posee propiedades similares,

además de una cierta compacidad espacial y, probablemente, identifique algún tipo de textura. De aquí surge la idea de que durante el proceso de clasificación no sólo se deben tener en cuenta las características individuales de los píxeles sino también su localización e información espacial. En este sentido, se han propuesto diferentes estrategias para incorporar la información espacial en los métodos de clasificación. Posteriormente, en la sección 2.2.3 se aborda específicamente este aspecto.

### **2.2.2 Características de las imágenes en la clasificación de texturas naturales**

Tal y como se ha mencionado previamente, por característica dentro de una imagen se entiende un elemento significativo de la misma, pudiendo ir desde una región formando parte de una determinada textura hasta un píxel o incluso agrupaciones de píxeles siguiendo un cierto patrón. En cualquier caso, cada característica se describe por una serie de propiedades que la identifican frente al resto. Las características y sus propiedades, constituyen las muestras para el proceso de clasificación, incluyendo la fase de entrenamiento y la decisión. En la literatura se han estudiado los comportamientos de diferentes características en la clasificación de las texturas (Valdovinos y col. 2005, Puig y García, 2006 o Hanmandlu y col. 2004). Existen dos enfoques: 1) métodos basados en las características espectrales de los píxeles (Puig y García, 2006; Rud y col., 2006; Hanmandlu y col., 2004; Yu y col., 2002; Kumar y col., 2001) y 2) métodos basados en regiones (Puig y García, 2006; Randen y Husøy, 1999; Wagner, 1999; Smith y Burns, 1997). Los métodos basados en los píxeles tratan de asignar el píxel a una clase determinada en función de sus propiedades específicas e individuales. Los métodos basados en las regiones tienen en cuenta las distribuciones espaciales de los píxeles dentro de la textura y por tanto su relación con otros píxeles en la región. En este último caso, cabe todavía distinguir entre texturas que presentan ciertas distribuciones regulares o patrones repetitivos, donde la unión de patrones define la textura global, y aquellas otras que carecen de distribuciones de este tipo, tratándose por tanto de texturas de naturaleza irregular.

En esta sección se analizan métodos basados en ambos enfoques. Así en la subsección 2.2.2.1 se revisan los métodos basados en los píxeles, bajo el punto de vista de su tratamiento individual. En la 2.2.2.2 se abordan los métodos que tienen en cuenta la distribución regular de las texturas, posponiendo el estudio de las texturas de

naturaleza irregular para la sección 2.2.3. La intención es tratarlas por separado, dado que constituyen la base fundamental del proceso de segmentación en las imágenes de los pinos, puesto que estas texturas irregulares son las que aparecen en las imágenes omnidireccionales utilizadas en los experimentos llevados a cabo mediante los métodos propuestos en la investigación. Finalmente, en la subsección 2.2.2.3 se abordan las técnicas que utilizan tanto los píxeles de forma individual como formando parte de regiones.

#### 2.2.2.1 Características espectrales a nivel de píxel

Como bien es sabido, los elementos básicos constituyentes de las imágenes son los píxeles, en este caso, considerados como características fundamentales de las propias imágenes, cuya propiedad más relevante es su naturaleza espectral. Existen técnicas de segmentación específicamente dedicadas a imágenes monocromas, es decir, imágenes de grises. No obstante, lo más frecuente es la utilización del color como atributo (Koschan y Abidi, 2008; Pajares y Cruz, 2007a; Du y col., 2004; Cheng y col., 2001).

El tratamiento del color ha recibido una importante atención en las últimas dos décadas, en su calidad de propiedad descriptora de los píxeles en las texturas (Jiji y Ganesan, 2008; Rud y col. 2006; Luo y col., 1998; Comaniciu y Meer, 1997; Pappas, 1992; Chang y col., 1994). Siguiendo esta línea, se han realizado diferentes estudios relacionados con los modelos de color, destacando los trabajos de Broek y Rikxoort (2004), Yu y col. (2002) o Kumar y col. (2001) donde se analizan, entre otros, los modelos de color *RGB* (*R*, red; *G*, green y *B*, blue), *HSI* (*H*, hue, *S*, saturation y *I*, intensity), *YIQ*, *YUV* o  $L^*u^*v^*$ , todos ellos encaminados hacia la clasificación de texturas. En general, destacan como los más eficientes los modelos *RGB* y *HSI*. El modelo *RGB* es muy utilizado en técnicas basadas en gráficos por computador y específicamente en Visión por Computador. Posee una gran dependencia de las características del computador donde se visualiza o representa el gráfico o la imagen. Por otro lado, el modelo *HSI* está más bien relacionado con el concepto de percepción humana según se indica en Tkalcic y Tasic (2003). En la literatura se han considerado otras representaciones de color tales como  $L^*a^*b^*$  o  $L^*u^*v^*$  debido también a su alto contenido perceptual (Yang y col., 2008).

En Trias-Sanz (2008) se realiza un estudio sobre diferentes modelos de color para la segmentación de imágenes de satélite correspondientes a entornos rurales, llegando a la conclusión de que en algunos casos resulta más conveniente utilizar ciertos modelos de color transformados, además del modelo *RGB* que también es utilizado.

Un análisis comparativo, con un elevado nivel de profundidad, sobre distintas representaciones del color se puede encontrar en Cheng y col. (2001), incluyendo las características monocromas. En dicho trabajo se concluye con una serie de ventajas e inconvenientes que se reproducen en la tabla 2.1.

El análisis de diferentes trabajos sobre segmentación de texturas en imágenes, lleva a la conclusión de que los modelos *RGB* y *HSI* compiten por ser utilizados en dicha tarea. Aunque en algunos casos puntuales con el modelo *RGB* se pueden obtener peores resultados que con el *HSI*, bien es cierto que aquél, en general, se presenta como un modelo satisfactoriamente aceptable, incluso a pesar de que sus tres componentes espectrales poseen una cierta dependencia lineal entre ellas (Comaniciu y Meer, 2002).

En algunos trabajos muy preliminares, como el de Robinson (1977), se llegó a la conclusión de que el modelo *HSI* era el más apropiado para la extracción de texturas en imágenes en color.

La tabla 2.1 recoge algunas ventajas e inconvenientes de los modelos de color desde el punto de vista de su utilización en el tratamiento de imágenes en general y de clasificación de texturas en particular, a modo de conclusión respecto de lo expuesto en las referencias anteriores.

Tabla 2.1 Ventajas e inconvenientes de los diferentes modelos de color

Modelo de color	Ventajas	Inconvenientes
<i>RGB</i>	Conveniente para dispositivos de visualización gráfica en computadores	Componentes espectrales correladas
<i>YIQ</i>	Útil para monitores de TV del sistema Americano	Persiste la correlación espectral debido a la transformación lineal desde el modelo <i>RGB</i>
<i>YUV</i>	Útil para monitores de TV del sistema Europeo	Persiste la correlación espectral debido a la transformación lineal desde el modelo <i>RGB</i>



Tabla 2.1 *Cont.*

<i>HSI</i>	Basado en el sistema de percepción humano. Útil en algunos casos donde la iluminación varía debido a que la componente <i>H</i> es invariante a ciertos tipos de sombras. Es útil si la iluminación varía en una imagen o entre imágenes sucesivas.	Presenta fuertes singularidades cuando en las ecuaciones de transformación el denominador se aproxima a cero, lo que ocurre con bastante frecuencia.  Es muy inestable con valores bajos de saturación.
Representaciones <i>CIE</i>	Se puede controlar la información del color y la intensidad independientemente.	Presenta fuertes singularidades debidas también a las ecuaciones de transformación desde el modelo <i>RGB</i>

Dentro del análisis del color existen otras aproximaciones relativas a la percepción visual, este es el caso de los trabajos propuestos por Vazquez y col. (2007) o Tkalcic y Tasic (2003). En ellos, a partir del histograma se identifican pequeñas variaciones del mismo con el fin de suprimirlas, consiguiendo así un cierto suavizado, y con ello la homogenización de las regiones de texturas.

Montiel y col. (2005) proponen utilizar la información proporcionada por las distribuciones de los niveles espectrales a través del análisis de los correspondientes histogramas de las imágenes que contienen las texturas.

Wangenheim y col. (2007), siguiendo la metodología basada en la división y fusión de regiones, aplican medidas de similitud para separar las clases que son diferentes y fusionar las que son similares. Las medidas de similitud se establecen en relación a las componentes de color.

#### 2.2.2.2 Características con patrones espaciales: regiones

Este tipo de características tienen la particularidad de presentar agrupaciones de píxeles con ciertas distribuciones regulares y espaciales de patrones. Existen diversas bases de imágenes con este tipo de texturas, tal es el caso del álbum de texturas de Brodatz (1966), disponible en numerosos sitios Web, o la base de datos VisTex, por citar sólo algunas. En la figura 2.2 se muestran varios ejemplos de tales texturas. Dentro de estos patrones se incluyen texturas procedentes de troncos de árboles, así como de hierbas, coincidiendo ambas con algunas de las analizadas en nuestras investigaciones,

razón por la cual se estudian estos patrones y algunos de los descriptores más utilizados en la literatura para su tratamiento.



Figura 2.2 Diferentes ejemplos de texturas procedentes de las bases de imágenes Brodatz y VisTex

Para el análisis de dichas texturas se utilizan diversos operadores de vecindad y filtros cuya descripción se puede encontrar en Pajares y Cruz, (2007a), Gonzalez y Woods (2008), Escalera (2001), Wagner (1999), Randen y Husøy (1999) o Smith y Burns (1997) entre otros. La idea consiste en situar una máscara alrededor de un determinado píxel, gracias a la cual se obtiene información sobre la distribución de los píxeles de la textura en la vecindad del píxel dado.

A continuación se ofrece una relación de diferentes métodos utilizados para la descripción de este tipo de texturas:

- Matrices de co-ocurrencia, con una serie de propiedades basadas en el histograma de las matrices, tales como: varianza, entropía o momentos de segundo orden entre otros (Chamorro-Martínez y col., 2007; Debeir y col., 2002).
- Filtros de Gabor (Wan y col., 2007; Palm y Lehmann, 2002; Dunn y Higgins, 1995; Porat y Zeevi, 1989).
- Máscaras de Laws (Gonzalez y Woods, 2008).

- Descomposición mediante la transformada de *wavelets* (Wang y Boesch, 2007; Pi y col., 2006; Permuter y col., 2006; Arivazhagan y Ganesan, 2003; Chang y Kuo, 1993; Unser, 1995).
- Descomposición en coeficientes de energía (Chen y col., 2005; Manjunath y Ma, 1996).
- Dominio de la frecuencia mediante análisis de Fourier (Lillo y col., 2007; Aiyer y col., 2005; Pyun y col., 2002; Clark y col., 1999).
- Transformada Discreta del Coseno (Li y col., 2006; Permuter y col., 2006).
- Patrones repetitivos mediante fractales (Tao y col., 2000).

En algunos casos se fusionan diversas estrategias, como ocurre en el método combinado descrito en Zhang y Wang (2004), que propone una estrategia multi-resolución mediante la descomposición de la imagen original en bandas de frecuencia a través de la transformada de *wavelets*, llegando a un cierto nivel de descomposición que es donde se utilizan descriptores basados en las matrices de co-ocurrencia (Pajares y Cruz, 2007a). En Chen y col. (2005) también se utiliza una estrategia multi-resolución para la descomposición de la imagen en diferentes bandas de frecuencia.

En Trias-Sanz y col. (2008) se han analizado diferentes tipos de descriptores de textura, tales como filtros de Gabor, patrones locales binarios, estimadores de Fourier o intensidad media, incluyendo algunos otros de naturaleza estructural. Las imágenes utilizadas son de satélite con predominio de zonas rurales. En este trabajo se concluye que el comportamiento de dichos descriptores no resulta apropiado para este tipo de imágenes.

En la figura 2.3 se reproduce por simplicidad la misma imagen de la figura 1.3(a), sobre la que pueden apreciarse alguna de las texturas del mismo tipo anteriormente mencionadas en relación a la figura 2.2, destacando el caso de los troncos, que constituyen un objetivo de interés de primera magnitud en nuestra investigación, si bien sobre este aspecto caben las siguientes matizaciones:

1. En la imagen de la figura 2.3 no aparecen patrones de textura repetitivos como en la figura 2.2, que permitan aplicar los descriptores que se usan para el tratamiento de esta última.
2. No existe una única textura, puesto que se distinguen claramente diversas zonas, tales como el cielo, hierba en el suelo, hojas en los árboles y los propios troncos de los árboles e incluso sobre estos últimos aparecen texturas con una importante variabilidad en la intensidad en función de la incidencia o no de los rayos del sol sobre ellos, a pesar de ser texturas de la misma naturaleza, es decir troncos de pinos.
3. Existe una clara distinción de los colores asociados a las bandas espectrales, lo que sin duda facilitará la identificación de determinadas texturas.



Figura 2.3 Ejemplo de imagen omnidireccional con diferentes texturas naturales

Diversos experimentos llevados a cabo sobre este tipo de imágenes, utilizando los descriptores mencionados anteriormente, no han producido los resultados que cabía esperar. Ello significa que las texturas presentes en nuestras imágenes no exhiben patrones de textura repetitivos, razón por la que se han descartado para su uso tanto en el proceso de segmentación como en el de correspondencia para capturar propiedades de los píxeles mediante dichos descriptores. Sin embargo, conectando con lo relativo a las componentes espectrales, el modelo de color *RGB* ha ofrecido resultados ciertamente satisfactorios, incluso el modelo *HSI* parece vislumbrarse como prometedor en futuras investigaciones tal y como se pone de manifiesto en los trabajos previos expuestos en el apéndice B. No obstante, en el segundo tipo de imágenes analizadas en nuestros

experimentos, procedentes de bosques de robles, cuya característica fundamental es su escasa iluminación, la utilización del color como propiedad relevante no proporciona resultados satisfactorios, ni en el modelo *RGB*, ni en el *HSI*, ni incluso en sus versiones realizadas. En este tipo de imágenes los descriptores de textura expuestos previamente tampoco producen resultados satisfactorios.

### 2.2.2.3 Combinación de características espectrales y espaciales

Las técnicas combinadas surgen bajo la idea de que la unión de esfuerzos resulta más productiva que los aportes individuales. En este caso, se trata de aprovechar la información conjunta proporcionada tanto por las características espectrales como la procedente de los patrones espaciales introducidos en la sección previa. Una herramienta muy utilizada para llevar a cabo esta fusión de información es la transformada de *wavelets*.

En la línea expuesta previamente se sitúa el trabajo de Chen y col. (2005). En él se utilizan dos tipos de características. Unas describen la composición local del color a nivel de píxel y las otras las características espaciales de textura en la imagen de intensidad. Cada tipo de característica se procesa por separado, para posteriormente proceder a su combinación. Las características de textura se obtienen descomponiendo la imagen en diferentes bandas de frecuencia, de forma que en cada nivel de la descomposición se utilizan aspectos relacionados con la frecuencia espacial.

En Wang y Boesch (2007) se combinan, tanto descriptores de color como de textura para la delimitación de masas forestales a partir de imágenes de satélite. Inicialmente la imagen se clasifica basándose exclusivamente en las componentes espectrales de color, donde cada píxel se etiqueta de acuerdo a la clase a la que pertenece, utilizando la técnica de división y fusión de regiones (Pajares y Cruz, 2007a; Gonzalez y Woods, 2008; Koschan y Abidi, 2008). Tras la clasificación inicial, se determinan los bordes de las regiones en base a las propiedades derivadas de los coeficientes de textura obtenidos mediante la transformada de *wavelets*.

En Wan y col. (2007) se combina el color y la textura utilizando los descriptores del estándar MPEG-7 para la extracción del contenido visual de las imágenes con texturas naturales.

Trias-Sanz y col. (2008) combinan de manera jerárquica las propiedades espectrales con los descriptores de texturas. En su estudio analizan diversos modelos de color combinados a partes iguales con diversos descriptores de texturas. En primer lugar obtienen una segmentación inicial mediante el color y seguidamente realizan una homogeneización de las regiones basándose en la distribución espacial de los descriptores de textura.

Ozden y Polat (2007) combinan dos tipos de propiedades: *a)* componentes espectrales de cada píxel y *b)* valores obtenidos a partir de los descriptores de textura. Todo ello considerando la información espacial proporcionada por la vecindad de cada píxel. La información procedente de los descriptores de textura la obtienen a través de los coeficientes de la transformada de *wavelets* (Pajares y Cruz, 2007a), considerando los coeficientes de detalle obtenidos mediante dicha transformada. Para incorporar la información espacial utilizan el procedimiento conocido como *Mean-Shift* (Comaniciu y Meer, 2002).

En Chanussot y col. (2006) se propone una estrategia que utiliza en un primer nivel la información espectral para clasificar imágenes urbanas obtenidas mediante el satélite IKONOS (SIC, 2010). En un segundo nivel utiliza la información espacial, que en este caso es información estructural, en el sentido de que aprovecha cierta información de este tipo para mejorar la clasificación. Exactamente, establece relaciones tales como que los tejados de las casas deben tener una forma geométrica determinada, las carreteras son estructuras alargadas, etc. De este modo, un píxel incorrectamente clasificado según alguna de las clases que identifican dichas formas se puede subsanar por la consideración de este tipo de información estructural.

### **2.2.3 Metodología de clasificación de texturas mediante información espacial**

Aunque el tema de la incorporación de la información espacial para la clasificación de texturas ha sido tratado previamente, en esta sección se enfoca desde una perspectiva más general, abordando así las texturas espacialmente irregulares, es decir, sin patrones predefinidos. En esto se diferencian de las texturas estudiadas en las secciones 2.2.2.2 y 2.2.2.3 donde los descriptores de textura se basan precisamente en distribuciones espaciales regulares.

Las relaciones espaciales se pueden establecer de diversas maneras. De entre todas ellas destaca el enfoque basado en los denominados *Campos Aleatorios de Markov* (*Markov Random Fields, MRFs*) (Geman y Geman, 1984; Li y col., 2000), que han sido ampliamente utilizados en la literatura. Por esta razón, en la sección 2.2.3.1 se realiza una revisión en profundidad de este enfoque. Las relaciones espaciales pueden establecerse también en función de ciertas propiedades estructurales, por ello en la sección 2.2.3.2 se consideran relaciones de esta naturaleza. Finalmente, en la sección 2.2.3.3 se revisan otros métodos, que establecen relaciones espaciales sin encuadrarse bajo una perspectiva concreta.

### 2.2.3.1 Información espacial mediante Campos Aleatorios de Markov

Básicamente, un *MRF* proporciona medidas de probabilidad sobre un dominio que posea relaciones espaciales. Este es el caso de las imágenes, ya que los objetos de la escena se proyectan en la imagen manteniendo sus relaciones espaciales. Esto es, los puntos vecinos en la escena 3D se proyectan también como píxeles vecinos en la imagen. El conjunto de posiciones donde se define un *MRF* se denomina *rejilla*. En las imágenes, la rejilla se genera a partir de los píxeles, que definen dichas localizaciones. Por consiguiente, bajo esta perspectiva, rejilla e imagen son entidades equiparables. Sobre cada localización espacial, definida por el correspondiente píxel de la imagen, se define una región de vecindad, que establece las relaciones entre las localizaciones que pertenecen a dicha región. A su vez, sobre la región de vecindad se define una función de energía local. La suma de todas las energías locales a lo largo de la rejilla, define una energía total. El objetivo de las técnicas basadas en los *MRFs* consiste en minimizar la energía total. Cada posición de la rejilla se caracteriza por poseer una probabilidad, que puede obtenerse a partir de la energía.

A continuación se proporciona un ejemplo ilustrativo sobre la aplicación de los *MRFs* al problema de la clasificación de texturas. En efecto, supongamos una región de vecindad sobre la rejilla de dimensión  $3 \times 3$  centrada en la localización  $i$  definida sobre el píxel del mismo nombre. A dicha región pertenecen las ocho localizaciones vecinas, entre ellas  $k$ . Los píxeles asociados con dichas localizaciones han sido clasificados previamente de forma que se les ha asignado una etiqueta de clase,  $l_i$  y  $l_k$  pertenecientes al conjunto de etiquetas posibles. Esto es, disponiendo de  $c$  clases, identificadas como

$w_h$  con  $h = 1, \dots, c$ , entonces, si el píxel  $i$  pertenece a la clase  $w_h$  la etiqueta  $l_i$  es  $h$ , y del mismo modo, si el píxel  $k$  pertenece a la clase  $w_j$  la etiqueta  $l_k$  es  $j$ .

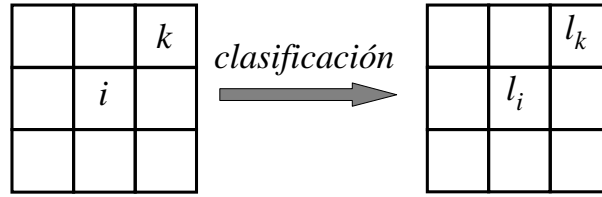


Figura 2.4 Relación de vecindad en un Campo Aleatorio de Markov

Se puede definir la función de energía total mediante la expresión siguiente:

$$E_L(l) = -\sum_L S(l_i, l_k) \text{ con } S(l_i, l_k) = \begin{cases} +1 & \text{si } l_i = l_k \\ -1 & \text{si } l_i \neq l_k \end{cases} \quad (2.1)$$

donde  $L$  se refiere a la región de vecindad alrededor de la localización  $i$ , siendo en este caso,  $L = 3 \times 3$ . A partir de la ecuación (2.1), se deduce que cuando ambas etiquetas tienen el mismo valor,  $S(l_i, l_k)$  toma el valor  $+1$ . Si todas las etiquetas, en la vecindad de  $i$ , toman el mismo valor que  $l_i$ , entonces la energía resulta ser  $E_L(l) = -8$ . En este caso lo que se expresa es que todos los píxeles de la vecindad han sido clasificados como pertenecientes a la misma clase que el píxel central. Por el contrario, si todas las etiquetas de las localizaciones, alrededor de la localización central, difieren de ésta, entonces  $S(l_i, l_k) = -1$  y  $E_L(l) = +8$ . Por tanto, en caso de compatibilidad en la clasificación de los píxeles en la vecindad, la energía es mínima, mientras que en caso de discrepancia total es máxima, siendo este uno de los planteamientos clave de los *MRFs*.

A partir de la energía local,  $E_L$ , se puede asignar a la localización  $i$  una cierta probabilidad de que la etiqueta  $l_i$  tome un valor concreto  $L$ . Por ejemplo, de acuerdo con la distribución de Gibbs se tendría la siguiente expresión (Haykin, 1994):

$$p(l_i = L) = Z^{-1} \exp \left\{ -\frac{1}{T} E_L(l) \right\} \quad (2.2)$$

$Z$  es una constante de normalización y  $T$  es otra constante análoga a la temperatura en procesos del tipo *Enfriamiento Simulado*, ambas pueden establecerse empíricamente. Bajo las suposiciones establecidas previamente, dada una localización  $i$  con etiqueta  $l_i$ ,



se deduce que a mayor compatibilidad entre  $i$  y sus vecinos, menor energía y por tanto mayor probabilidad asignada a la localización  $i$ . En caso de discrepancia entre  $i$  y sus vecinos, ocurre lo contrario, esto es, mayor energía y menor probabilidad.

Las diferencias entre los métodos que se citan a continuación, donde los *MRFs* constituyen su fundamento, estriban en el cómputo de esta probabilidad y naturalmente en cómo se combina esta probabilidad durante la clasificación según el método utilizado.

En Deng y Clausi (2004, 2005a,b) la energía consta de dos términos. Un término se obtiene a partir de las relaciones entre los píxeles por medio de sus características espectrales, de forma que la compatibilidad en la vecindad se obtiene según las similitudes de las características espectrales entre el píxel central y sus vecinos. El segundo término se obtiene a partir de las etiquetas de clase de una forma similar a la descrita previamente. Con ellos se calculan dos valores de probabilidad, que se combinan mediante el producto de ambos, obteniendo finalmente un único valor de probabilidad. El proceso de actualización de las probabilidades resultantes del producto se lleva a cabo mediante un procedimiento iterativo hasta conseguir el mínimo valor posible de energía.

El trabajo de Xia y col. (2006) sigue un enfoque similar al expresado previamente en los trabajos de Deng y Clausi (2004, 2005a,b), si bien aborda específicamente un problema importante que surge en los bordes de las regiones donde las texturas rompen su homogeneidad como consecuencia del tránsito de una región a otra. En este caso, en lugar de extraer primero las características de las regiones para proceder a su clasificación posterior, que suele ser una práctica general, se alterna sucesivamente entre obtención de características y clasificación. No obstante, el método también calcula dos términos de energía, a partir de los que obtiene las probabilidades en cada localización, que son combinadas mediante el producto. La actualización de las probabilidades se lleva a cabo mediante el conocido procedimiento de *Enfriamiento Simulado* (Duda y col, 2000). Un esquema similar al de Xia y col. (2006) es el propuesto por Borges y col. (2007).

En la línea de los trabajos anteriores se encuentran las investigaciones de Cariou y Chehdi (2008) o Allili y Ziou (2007), donde se plantean métodos basados en multi-

resolución y por tanto en la descomposición de la imagen en varios niveles. En el primer nivel la imagen se divide en regiones regularmente espaciadas con una dimensión de ventana fija. A partir de dichas regiones se estima: *a)* el número de clases de texturas; *b)* las características que describen las texturas. En el segundo nivel se obtiene la energía asociada a las localizaciones espaciales, bajo el paradigma de los *MRFs*, de una forma similar al procedimiento descrito en el párrafo previo. El proceso de actualización de las probabilidades se realiza mediante el método conocido en terminología inglesa como *Expectation Maximization (EM)* (Duda y col., 2000).

En Pyun y col. (2007) se propone una estrategia en la que se estiman las funciones de densidad de probabilidad suponiendo distribuciones Gaussianas, lo que constituye una práctica común en muchos métodos de clasificación. La estimación de dichas funciones se lleva a cabo utilizando un método de cuantización vectorial (*LVQ*) en lugar de utilizar el clásico *EM*. Su justificación se fundamenta en una convergencia más rápida que la requerida por el *EM*. Como en el trabajo previo, la información espacial se introduce considerando la vecindad bajo el paradigma de los *MRFs*.

Huang (1999) y Campadelli y col. (1997) extraen la información espacial basándose en la similitud de etiquetas con las que han sido clasificados los píxeles vecinos en relación al central. Se aplica un proceso de minimización de la energía mediante la red neuronal de Hopfield. El cálculo de la energía se obtiene mediante un procedimiento idéntico al descrito en la ecuación (2.1).

Algunas otras investigaciones en el ámbito de los *MRFs*, con planteamientos similares a los anteriores, son los trabajos de Pyun y col. (2007), Yang y Jiang (2003), Melas y Wilson (2002), Li y col. (2000), Barker y Rayner (2000), Sarkar y col. (2000), Andrey y Tarroux (1998), Krishnamachari y Chellapa (1997), Manjunath y col. (1990).

Desde el punto de vista de nuestra trabajo de investigación y en particular en el caso de los pinos, el objetivo consiste por un lado en excluir ciertas texturas del proceso de correspondencia (hierba en el suelo, cielo y hojas de los árboles) y por otro lado en identificar las texturas pertenecientes a los troncos por su interés en los inventarios forestales. Pensando en la aplicación de los *MRFs* a nuestro problema, nos encontramos que sólo en el caso de las texturas correspondientes a la zona de cielo tendría ciertas posibilidades de éxito, ya que en el caso de los troncos o hierba del suelo aparecen

abundantes zonas de sombra y variaciones de intensidad, que impiden la aplicación del concepto de propiedades similares en la vecindad. En el caso de las texturas pertenecientes a las hojas de los árboles el problema se revela si cabe más acusado, puesto que en ellas la homogeneidad de la textura brilla por su ausencia. Si nos fijamos en las imágenes procedentes de los bosques de robles la problemática se manifiesta en sentido contrario, ya que en este caso existe una clara confusión entre la parte inferior de los troncos y las texturas del suelo, ambas presentan propiedades similares de intensidad, con lo que la aplicación de los *MRFs* conduce a la identificación de toda esa masa de vegetación como una única textura, lo cual no resulta deseable.

### 2.2.3.2 Relaciones estructurales

El trabajo de Chen y col. (2005) incorpora la información espacial mediante la inyección de conocimiento procedente de la percepción humana. Se parte de los niveles elementales de color, a nivel de píxel (proceso local), para posteriormente utilizar descriptores de textura que incorporan información semántica. Un ejemplo de este tipo de información se refiere por ejemplo a situaciones tales como que en el caso de nuestras imágenes circulares, las texturas correspondientes al cielo se sitúan preferentemente en la parte central de la imagen y las del suelo en los bordes de la circunferencia.

En el ámbito de las imágenes de satélite, la información espacial se ha incorporado también aplicando un conocimiento de índole estructural, lo que supone la aplicación de inteligencia de nivel superior. Los píxeles de una determinada estructura están rodeados de píxeles de esa misma estructura o de estructuras compatibles, por ejemplo en una zona urbana no es probable encontrar un píxel asociado con una zona de cultivo perteneciente al ámbito rural. En esta línea se encuentran los trabajos de Barnsley y Barr (1996) o Gong y Howarth (1990).

### 2.2.3.3 Otros enfoques

En este apartado se referencian varios métodos, que sólo tienen en común el hecho de la incorporación de información espacial, lo cual se lleva a cabo bajo distintas perspectivas, que a su vez difieren de las expresadas anteriormente. Algunos de ellos se han utilizado en otros contextos diferentes al de las imágenes de texturas en entornos de

exterior, sin embargo se incluyen aquí como ejemplos adicionales para la incorporación de la información espacial.

El método de Xie y Tsui (2004) se basa en sustituir el procedimiento de actualización iterativo *EM* por el clásico método conocido como Agrupamiento Borroso. Este último algoritmo se describe en el capítulo tres, sección 3.2.3 por ser uno de los métodos base utilizados en el presente trabajo de investigación. La aportación del trabajo de Xie y Tsui (2004) consiste en la definición de una nueva métrica, que incorpora tanto la información de similitud entre las características espectrales, como la información espacial. Esta última se obtiene mediante el cómputo de los grados de pertenencia de los píxeles vecinos, respecto de un píxel dado, a sus correspondientes clases. Cuanto mayor sea el grado de pertenencia de los vecinos, mayor es la contribución de éstos a la minimización de la función objetivo.

En la misma línea del anterior se encuentra la propuesta de Liew y col. (2003). En su trabajo, propusieron considerar los ocho píxeles vecinos al píxel central que se está tratando de clasificar; si los vecinos tienen valores similares y altos de pertenencia a la misma clase que el píxel central, el valor del grado de pertenencia de éste se modifica según la tendencia de sus vecinos. Podría considerarse como una especie de suavizado en función de la vecindad, por lo que su efecto sobre la modificación de los grados de pertenencia de píxeles que han sido incorrectamente clasificados tiene un efecto importante. Este método trata bastante bien las imágenes ruidosas de cara a una posible homogeneización de las texturas.

El trabajo propuesto por Krishnapuram y Freg (1992) incluye información de las coordenadas en el espacio 3D. Cuando dicha información está disponible, resulta de gran ayuda para la clasificación, ya que los puntos situados en localizaciones espaciales contiguas tenderán a agruparse en localizaciones vecinas de la imagen bidimensional. Su principal inconveniente estriba en que se necesita tener disponible la información de las coordenadas espaciales con anterioridad al proceso de clasificación.

Kupinski y Gigger (1998), en el campo de las imágenes médicas, utilizaron una función isotrópica Gaussiana para interpretar el conocimiento a priori de cada forma de la región objetivo. La información espacial de la imagen se multiplica por esta función Gaussiana para eliminar valores de píxeles distantes, de tal forma que la imagen es

posteriormente segmentada utilizando métodos de umbralización o técnicas de crecimiento de regiones (Pajares y Cruz, 2007a). La información espacial incide en la mejora de las técnicas de umbralización o crecimiento de regiones. Desafortunadamente, a pesar de los buenos resultados mostrados, la generalización de este método no resulta fácil, particularmente en las imágenes omnidireccionales utilizadas en nuestro trabajo, donde las texturas fuera de interés (hierba en el suelo, cielo y hojas en los árboles) no poseen formas determinadas dada su naturaleza. Por consiguiente la utilización de las semillas, como elementos clave en las técnicas de crecimiento de regiones, supone una dificultad añadida.

El uso de los contornos activos (*Active Contour Model, ACM*), es decir, la delimitación de las regiones mediante contornos, ha sido otra de las técnicas utilizadas en la segmentación de las imágenes con texturas. El problema que entrañan es que necesitan un conocimiento previo de los objetos que se quieren identificar. Es evidente que en el caso de las texturas naturales esta suposición no es posible debido a que no existen objetos delimitados o con formas específicas, y además las formas de las texturas difieren considerablemente de unas imágenes a otras. Por otro lado, requieren un elevado coste computacional, a pesar de las propuestas basadas en B-splines, que tratan de minimizar dicho coste (Precioso y col., 2005). Otras propuestas, que intentan reducir el coste computacional, se encaminan hacia la segmentación de imágenes en varios niveles, de forma que en un nivel de resolución bajo se obtienen los primeros resultados que guían el proceso en los niveles de más alta resolución (Kim y col., 2006).

## 2.3 Correspondencia estereoscópica

### 2.3.1 Restricciones aplicadas en correspondencia estéreo

En correspondencia estereoscópica generalmente se aplican un conjunto de restricciones para resolver el problema general de la correspondencia, éstas son: *epipolar*, *semejanza*, *unicidad*, *orden posicional*, *continuidad de la disparidad* y *relaciones estructurales*. Estas restricciones se definen de forma general como sigue (Pajares y Cruz, 2007a; Scharstein y Szeliski, 2002):

- *Epipolar*: las imágenes de una misma entidad 3D deben proyectarse sobre la misma línea epipolar. Esta restricción se deriva de la geometría del sistema y de

la proyección utilizada. Tanto en los sistemas convencionales como en el utilizado en este trabajo se requiere que las cámaras estén alineadas, tal y como se describe en el capítulo cuatro. Un método para lograr el alineamiento en los sistemas convencionales se puede encontrar en Reid y Beardsley (1996).

- *Semejanza*: las dos imágenes de la misma entidad 3D deben tener propiedades o atributos similares.
- *Unicidad*: para cada característica en una imagen debe haber una única característica en la otra imagen, salvo que se produzca una oclusión y no haya correspondencia de alguna característica.
- *Orden posicional*: dadas dos características en una determinada imagen, por ejemplo la izquierda, situada una a la derecha de la otra, esta restricción supone que este mismo orden se mantiene en la imagen derecha para sus respectivas características homólogas.
- *Continuidad de la disparidad*: asume que las variaciones de disparidad en la imagen son generalmente suaves, es decir que si consideramos un mapa de disparidad éste se presenta continuo salvo en unas pocas discontinuidades. Este principio también aparece bajo distintas formas y a veces con alguna pequeña variación, tal es el caso de *Disparidad Diferencial Mínima* en Medioni y Nevatia (1985) o Pajares y col. (1998) o *Continuidad "figural"* en Pollard y col. (1981).
- *Relaciones estructurales*: supone que los objetos están formados por aristas, vértices o superficies con una cierta estructura y una disposición geométrica entre dichos elementos.

En nuestro primer enfoque orientado a resolver el problema de la correspondencia en el conjunto de pares de imágenes omnidireccionales procedentes de los bosques de pinos que presentan alta variabilidad de intensidad, la correspondencia se establece a nivel de píxel. Los píxeles bajo correspondencia deben satisfacer una serie de restricciones que por la naturaleza de las imágenes son: *epipolar*, *semejanza*, *unicidad* y *continuidad en la disparidad*. A continuación se explica cómo se aplican estas restricciones en las estrategias propuestas:

- *Epipolar*: derivada de la geometría del sistema, dado un píxel en una imagen su homólogo en la otra imagen del par estereoscópico se sitúa sobre la línea epipolar cuyas ecuaciones se obtienen a partir de la geometría del sistema y la proyección basada en la óptica asociada a las lentes de ojo de pez.
- *Semejanza*: píxeles homólogos presentan atributos o propiedades similares.
- *Unicidad*: un píxel en una imagen debe corresponderse con un único píxel en la otra imagen, excepto si se producen oclusiones.
- *Continuidad de la disparidad*: los valores de disparidad en la vecindad cambian suavemente, excepto en unas pocas discontinuidades tales como los bordes de los troncos de los árboles.

Nuestro segundo enfoque está orientado a resolver el problema de la correspondencia en imágenes procedentes de bosques de robles con pobre iluminación y alto contraste entre el cielo y los troncos en la parte central. A diferencia del enfoque anterior, aquí la correspondencia se establece a nivel de regiones. Estas correspondencias deben satisfacer una serie de restricciones que en este caso son *epipolar*, *semejanza*, *orden posicional* y *unicidad*:

- *Epipolar*: dada una región en una imagen, la región correspondiente en la segunda imagen debe encontrarse siguiendo la llamada línea epipolar.
- *Semejanza*: regiones homólogas deben poseer atributos o propiedades similares.
- *Orden posicional*: la posición relativa entre dos regiones en una imagen generalmente se preserva en la otra imagen para sus regiones homólogas.
- *Unicidad*: cada región en una imagen debe corresponderse con una única región en la otra imagen, excepto si se producen oclusiones.

### 2.3.2 Técnicas en correspondencia estereoscópica

La revisión del estado del arte en visión estereoscópica nos permite distinguir dos tipos de técnicas ampliamente utilizadas en esta disciplina: *basadas en el área* (*area-based*) y *basadas en las características* (*feature-based*).

Las técnicas basadas en el área utilizan patrones de intensidad en la vecindad local de un píxel en una imagen con patrones también de intensidad en una vecindad homóloga de un píxel en la otra imagen del par estereoscópico (Cochran y Medioni, 1992; Scharstein y Szeliski, 2002). Dos tipos de enfoques se sitúan bajo esta categoría. El primero hace referencia al coeficiente de correlación y el segundo a las medidas estadísticas, usadas generalmente también para identificar texturas.

Los métodos basados en las características (Tang y col., 2002) calculan atributos para esas características (en nuestro caso, píxeles o regiones) bajo correspondencia; pueden ser simples atributos, tales como el color de los píxeles (Klaus y col., 2006) o propiedades obtenidas aplicando algún operador como el módulo del gradiente (Klaus y col., 2006) y el ángulo del gradiente o la Laplaciana (Lew y col., 1994). Ello a pesar de que en algunos contextos tanto el gradiente como la Laplaciana podían llegar a presentar cierta sensibilidad al ruido. En realidad, estos operadores tienen en cuenta los píxeles y sus vecinos; por tanto, desde este punto de vista podrían considerarse como basados en el área. El color es otro de los atributos que puede utilizarse a nivel de píxel individual o con la intervención de un entorno de vecindad (Klaus y col., 2006).

Los métodos basados en las características utilizan normalmente conjuntos de píxeles con atributos similares, ya sean píxeles pertenecientes a bordes (Tang y col., 2002; Grimson, 1985; Ruichek y Postaire, 1996), los bordes correspondientes a sí mismos (Medioni y Nevatia, 1985; Pajares y Cruz, 2006; Scaramuzza y col., 2008), regiones (McKinnon y Baltes, 2004; Marapane y Trivedi, 1989) o enfoques jerárquicos (Wei y Quan, 2004) donde primero se establece la correspondencia entre bordes o esquinas y después las regiones. En Tang y col. (2002) se utilizan regiones con los tres siguientes atributos específicos para correspondencia: área, centroide y ángulos. Éstos son utilizados en el tratamiento de las imágenes procedentes de los bosques de robles por su alta capacidad para describir las regiones que se utilizan como características en este tipo de imágenes.

Existen numerosos trabajos en los que se utilizan los atributos anteriores para correspondencia mediante la aplicación de la restricción de semejanza. En Chehata y col. (2003) dichos atributos se materializan en los siguientes: área, *bounding box*, momentos estadísticos espaciales. En Kaick y Mori (2006), aunque bajo un contexto de clasificación, se utilizan los momentos estadísticos de primer y segundo orden en el



espacio de color *HSI*; estos atributos se obtienen a partir de los histogramas. En Renninger y Malik (2004) también se aplican descriptores de texturas, tales como los bancos jerárquicos de filtros. En Hu y Yang (2008) y Premaratne y Safaei (2008) se han aplicado satisfactoriamente momentos invariantes, donde se concluye sobre la conveniencia de su uso en la correspondencia basada en regiones por la mejora en la exactitud a la hora de obtener las disparidades. Esto constituye un elemento motivador para el caso de las imágenes de robles analizadas en este trabajo. En López y Pla (2000) se propone un método basado en grafos para tratar con errores de segmentación en correspondencia basada en regiones. Los nodos en el grafo son los posibles pares potenciales de correspondencias, mientras que a los arcos se les asigna valores teniendo en cuenta una medida de similitud entre las regiones bajo correspondencia.

En Wang y Zheng (2008) se extraen las regiones mediante un algoritmo de segmentación basado en el color y los píxeles pertenecientes a las regiones son emparejados obteniéndose un mapa de disparidad, que luego es refinado aplicando optimización cooperativa mediante el ajuste de algunos parámetros en las disparidades de las regiones segmentadas. En Ansari y col. (2007) el color también se emplea para segmentar las regiones.

En Scaramuzza y col. (2008) se utilizan líneas verticales como características en imágenes omnidireccionales y se calcula un descriptor invariante a rotaciones. Esta invarianza a la rotación también resulta útil en nuestras imágenes de robles, como veremos posteriormente en el capítulo cuatro.

Ya hemos mencionado reiteradamente, que nuestro interés se centra en determinar las correspondencias de los píxeles o regiones asociados a los troncos de los árboles porque son las estructuras determinantes en el análisis forestal. Por lo tanto, una vez clarificado que la aplicación de las restricciones estereoscópicas resulta una cuestión de vital importancia, el tema ahora se centra en decidir si se utilizan técnicas basadas en el área o en las características, que constituyen lo que en la bibliografía especializada se conocen como *area-based* o *feature-based* respectivamente. A continuación se proporcionan algunos detalles sobre la conveniencia o no de su uso en función de qué circunstancias.

*Técnicas basadas en el área (area-based)*

- 1) La correspondencia se lleva a cabo píxel a píxel siguiendo las líneas epipolares. Esto no requiere de conocimiento previo sobre si el píxel bajo análisis pertenece a un tronco o no.
- 2) La correspondencia se establece por similitud entre las propiedades de los píxeles bajo análisis. El principal inconveniente es que en nuestro segundo grupo de imágenes, con pobre iluminación y alto contraste, los troncos y la hierba en el suelo presentan firmas espectrales similares. Las dos presentan una tonalidad con tendencia al gris oscuro para las imágenes en la figura 1.5 y verde para las imágenes de la figura 1.6. Por lo tanto, en las partes comunes donde el suelo y los troncos se confunden, la identificación de éstos se convierte en una tarea ardua y difícil.
- 3) La parte de la imagen asociada al cielo resulta ser muy homogénea, lo que hace que la correspondencia píxel a píxel también se convierta en una difícil tarea.
- 4) Debido a las dificultades anteriores, si se lleva a cabo la correspondencia píxel a píxel, tras la obtención del mapa de disparidad se hace necesario identificar los píxeles pertenecientes a los troncos por similitud de disparidades.

*Técnicas basadas en las características (feature-based)*

- 1) Es la opción natural que un experto humano aplica de forma intuitiva. En efecto, la correspondencia que cualquiera de nosotros realiza ante un problema de este tipo supone comparar árbol a árbol las regiones que los representan en ambas imágenes.
- 2) Lo anterior implica que el humano analiza los árboles y por tanto los troncos aplicando similitudes de forma entre ellos y considerando también su ubicación en la imagen basada en la restricción epipolar proporcionada por la geometría del dispositivo estereoscópico de medida. La restricción de orden posicional resulta de gran ayuda para tomar la decisión final.
- 3) La orientación radial de los troncos hacia el centro óptico en las imágenes podría aprovecharse para establecer la correspondencia, aunque como se ha

mencionado previamente no precisamente como un elemento decisor en sí mismo, entre otras razones porque los árboles no crecen completamente verticales.

- 4) El principal inconveniente del enfoque basado en las características en nuestro problema específico, desde el punto de vista de la automatización del proceso, es que los troncos deben ser identificados previamente y a continuación extraer un conjunto de propiedades para su identificación.

Como puede fácilmente deducirse, cada método posee sus ventajas e inconvenientes que conviene valorar en el momento de su aplicación. Una conclusión importante concerniente a ambos métodos es que es muy importante que los píxeles pertenecientes a los troncos puedan aislarse e identificarse unívocamente. Bajo esta perspectiva y basándonos en la observación de las imágenes realzadas pertenecientes al segundo grupo de imágenes correspondientes a los robles, figura 1.6, hemos intentado aplicar métodos de identificación de texturas para segmentar los árboles bajo diferentes espacios de color siguiendo las ideas expuestas en el trabajo de Trias-Sanz y col. (2008). Los espacios de color investigados fueron los clásicos *RGB*, *HSI*, *CIE XYZ*, *L\*a\*b\** y *L\*u\*v\**, también la log-cromaticidad oponente (*log-opponent chromaticity*) y adicionalmente la cromaticidad rojo/cyan propuesta en Trias-Sanz y col. (2008). Desde el punto de vista de las texturas y basándonos en los resultados obtenidos por los diferentes espacios de color, hemos aplicado técnicas basadas en descriptores estadísticos tales como la varianza y la intensidad promediada, ambas investigadas en Gonzalez y Woods (2008), así como también los filtros de Gabor (Bandzi y col., 2007).

En Ansari y col. (2007), se utiliza un método basado en una función de coste para la correspondencia obteniendo diferencias de color entre las características, generalmente píxeles, en ambas imágenes del par estereoscópico, básicamente para desambiguar falsas correspondencias. El problema principal en nuestras imágenes es relativo a la similitud entre el color y las texturas en los troncos y en el suelo. Después de varios experimentos, no hemos logrado de manera satisfactoria la separación de estos dos tipos de texturas a través de los enfoques anteriores. Esto significa que los troncos no son separables fácilmente a través de estos métodos. Además, como se ha mencionado previamente, en López y Pla (2000) se propone un método de correspondencia basado en grafos para fusionar y separar regiones que han sido

segmentadas incorrectamente, esto facilita el proceso de correspondencia posterior en el caso que nos compete, si bien en el caso particular de las imágenes analizadas en la presente investigación, esta técnica no ha proporcionado los resultados esperados por las características específicas de las propias imágenes. En efecto, como se mencionó previamente, en nuestro segundo grupo de imágenes de robles, el problema principal es la separación de los troncos del suelo en lugar de la fusión de esas regiones. Cuando intentamos identificar los troncos, siempre obtenemos una amplia región única limitando con la circunferencia exterior de la imagen válida; la división de esta amplia región, basada en las diferencias de intensidad o textura, se convierte en una tarea ciertamente difícil. Sin embargo, para este tipo de imágenes podemos aplicar el concepto de división, si bien bajo consideraciones geométricas, como se verá detalladamente en el capítulo tres y con resultados ciertamente satisfactorios.

El método propuesto en Wang y Zheng (2008) requiere del cálculo previo de un mapa inicial de disparidad denso, que es posteriormente refinado ajustando un plano sobre las regiones segmentadas. En nuestro enfoque, un problema importante es el relativo a la segmentación de las regiones y precisamente el cálculo de ese mapa de disparidad, lo que en sí mismo constituye un fin y no un medio, por lo que no resulta útil en nuestro análisis.

## **2.4 Resumen y conclusiones del análisis bibliográfico**

A la vista de la revisión bibliográfica realizada en conjunción con el estudio de la problemática existente en relación al análisis de los dos tipos de imágenes utilizadas, se concluye que las imágenes procedentes del bosque de pinos se adaptan mejor para su tratamiento a nivel de píxel, mientras que las imágenes del bosque de robles admiten un tratamiento más específico desde la perspectiva de las regiones. Por estas razones, planteamos sendos enfoques desde ambos puntos de vista, esto es: píxeles y regiones. La idea consiste en el diseño de una estrategia general para cada enfoque con la perspectiva de su aplicación en el futuro a otros tipos de imágenes procedentes de otros entornos forestales, que si bien las imágenes pueden diferir puntualmente, de forma general plantean una problemática similar. El objetivo consistiría en la adaptación de las estrategias aquí propuestas a los nuevos entornos con el mínimo esfuerzo en el futuro.

### 2.4.1 Proceso completo basado en píxeles

Este proceso, uniendo segmentación y correspondencia, resulta útil para las imágenes que presentan una alta variabilidad de intensidad como consecuencia de la incidencia de los rayos solares sobre las estructuras que definen las texturas en la imagen y que además no presentan patrones repetitivos de textura.

En la estrategia que se presenta, el modelo de color *RGB* captura convenientemente la mencionada variabilidad, por lo que se ha elegido como el más apropiado en este momento, todo ello sin perjuicio de las supuestas mejoras que puedan surgir en este sentido como consecuencia de la utilización de otros modelos en el futuro, como se apunta en el apéndice B.

El proceso de segmentación se diseña como sigue. Se identifican tres tipos de texturas (hierba en el suelo, cielo y hojas en los árboles) mediante una estrategia de reconocimiento de patrones basada en la aplicación de dos clasificadores, a saber:

- Agrupamiento Borroso (*FC*).
- Estimador Paramétrico Bayesiano (*PB*).

Se eligen estos clasificadores por su simplicidad y probada eficacia en diversas aplicaciones relativas a la clasificación de texturas naturales, particularmente cuando ambos se utilizan de forma combinada, tal y como se ha puesto de manifiesto en los trabajos expuestos en Guijarro (2009) y Pajares y col. (2009). Bien es cierto que se podrían haber elegido otros o diferentes estrategias para su combinación, sin embargo, dada la envergadura del problema tratado, se ha optado por acotar la problemática comenzando por este planteamiento.

La identificación de las texturas a excluir del proceso de correspondencia, nos permite afrontar éste con mayores garantías de éxito. A tenor de lo cual, el proceso de correspondencia se diseña como sigue. Sin pérdida de generalidad, dado un píxel en la imagen izquierda, se aplica la restricción epipolar para determinar una lista de candidatos, los cuales son potenciales correspondencias en la imagen derecha. Cada candidato se convierte en una alternativa para el primer píxel. Para cada par de píxeles se aplica la restricción de semejanza basándose en los seis atributos siguientes: *a)*

coeficiente de correlación, *b*) color para cada píxel, *c*) varianza como una medida de la textura, *d*) módulo del gradiente, *e*) ángulo del gradiente y *f*) Laplaciana. El gradiente se obtiene a través del operador de Sobel, aunque se podrían utilizar otros operadores de borde.

Basándose en los seis atributos se calculan seis medidas de semejanza entre cada píxel y los píxeles de la lista de candidatos. Estas medidas de semejanza se combinan convenientemente. La decisión final sobre quién es el candidato seleccionado como pareja entre los disponibles en la lista, se toma según el máximo grado de correspondencia recibido por éste en función de la aplicación de los siguientes paradigmas:

- La Integral Fuzzy de Choquet (*CFI*).
- La Integral Fuzzy de Sugeno (*SFI*).
- La Teoría Dempster-Shafer (*DES*).
- La Toma de Decisiones Multicriterio bajo la perspectiva de la lógica Fuzzy (*MCDMF*).

Los dos primeros son técnicas utilizada para combinar clasificadores (Kuncheva, 2004) y han sido convenientemente adaptadas en este trabajo para ser aplicadas en correspondencia. La selección de un único candidato supone la aplicación de la restricción de unicidad.

En resumen, en este momento se han aplicado tres restricciones en correspondencia estereoscópica (epipolar, semejanza y unicidad). El mismo proceso se repite tomando como referencia la imagen derecha, buscando ahora emparejamientos en la izquierda. El objetivo consiste en deshacer, hasta donde sea posible, las ambigüedades surgidas en los emparejamientos, a la vez que se tiende hacia decisiones más clarificadoras. En efecto, si dos píxeles se corresponden igualmente en ambos sentidos izquierda-derecha, esto constituye un signo positivo de correspondencia correcta.

Una vez emparejados dos píxeles, se determina su disparidad en términos de su diferencia en grados en función de sus respectivos ángulos en coordenadas polares, tal y

como se describe en el capítulo cuatro. Con ello se obtiene un mapa de disparidad, tomando como referencia la localización de cada píxel en la imagen izquierda, para cada par de imágenes en términos de esas diferencias de grado.

El siguiente objetivo consiste en mejorar el mapa de disparidad hasta donde sea posible. El objetivo es eliminar posiciones con valores erróneos en las disparidades y suavizar las áreas que corresponden a los troncos situados a una distancia única, lo que debería llevar a valores de disparidad similares. Esto supone exactamente la aplicación de la restricción de continuidad de la disparidad y de algún modo la filosofía derivada de los *MRFs*, anteriormente referidos, si bien bajo un contexto diferente. El suavizado se lleva a cabo aplicando una estrategia que considera no sólo los valores de disparidad aislados en la posición de cada píxel, sino los píxeles vecinos. Con tal propósito se han seleccionado los paradigmas enumerados a continuación, descritos en el capítulo cuatro, recurriendo para su aplicación a los principios de la escuela de psicología de la Gestalt (Koffka, 1935; Wang, 2005), englobando así de paso el esquema de los *MRFs*. Dichos paradigmas son concretamente:

- Los Mapas Cognitivos Fuzzy (*FCMs*).
- La red Neuronal de Hopfield (*HNN*).

La elección de los *FCMs* se ha debido a su éxito en otras propuestas de correspondencia estereoscópica aunque en un contexto diferente y bajo criterios también diferentes (Pajares y Cruz, 2006).

La *HNN* se ha seleccionado por el hecho de tratarse de una estrategia de optimización, cuyo proceso puede ser controlado por minimización de una función de energía, que se establece a tal efecto. Además, la *HNN* también ha sido utilizada satisfactoriamente en otras propuestas de correspondencia estereoscópica, aunque en un contexto diferente y bajo diferentes criterios (Pajares y col., 1998).

Los detalles relativos al proceso de segmentación basado en píxeles se proporcionan en la sección 3.2, mientras que los relativos al proceso de correspondencia bajo el mismo enfoque se describen en la sección 4.2.

### 2.4.2 Proceso completo basado en regiones

Para el tipo de imágenes procedentes del bosque de robles con sus características inherentes relativas a la escasez de iluminación y a la vez relativo alto contraste en su parte central, se ha diseñado una estrategia basada en regiones que aborda el problema tratando de discernir entre las texturas de los troncos respecto de las del suelo a pesar de su similitud en las imágenes. De este modo, la parte de la imagen donde las texturas pertenecen a los troncos pueden distinguirse fácilmente de aquellas pertenecientes al cielo, separándolas mediante diferencias de contraste en la intensidad allí donde es posible, mientras que en la parte donde las texturas pertenecientes a los troncos y al suelo son indistinguibles se estima la posición del tronco guiándose por el conocimiento geométrico derivado del proceso de discriminación previo.

Tras la extracción de regiones identificativas de los troncos se obtiene un conjunto de propiedades para cada región, que de algún modo describen el tronco representado por su correspondiente región. Posteriormente, la aplicación de las restricciones estereoscópicas, involucrando similitudes entre propiedades y relaciones geométricas basadas en la geometría del sistema, permite establecer la correspondencia de los troncos en ambas imágenes del par estereoscópico.

En resumen, el proceso completo de correspondencia estereoscópica implica dos pasos principales, a saber: la segmentación del tronco y la correspondencia de los troncos. Como en el caso de la estrategia anterior basada en píxeles, ambos procesos constituyen un todo inseparable.

### 2.4.3 Resumen global

Volviendo sobre la figura 2.1 y a modo de resumen, se acaban de perfilar sendos procesos de segmentación y correspondencia, que se perfilan como dos componentes esenciales en la arquitectura global del sistema mostrado en dicha figura.





# 3

## Segmentación

### 3.1 Introducción

En los capítulos precedentes se expuso la etapa de segmentación como una de las fases clave en el proceso global de la visión estereoscópica. Se trata de la etapa comprendida entre la captura de imágenes, en la que se incluye la geometría del sistema, y el proceso de correspondencia, tal y como quedó reflejado en el diagrama de la figura 2.1.

Por otro lado, también se analizaron dos clases de imágenes con peculiaridades bien diferentes, pertenecientes a sendos tipos de bosques, pinares y robles. Volviendo sobre dicho estudio, respecto del primer tipo, se convino en que las características más apropiadas para llevar a cabo el subsiguiente proceso de correspondencia estereoscópica eran los píxeles, por el conjunto de razones expuestas a lo largo de los temas precedentes, y por entender que el análisis de las imágenes basado en cualquier otro tipo de características conlleva muchas más dificultades que ventajas. Por el contrario, respecto del segundo tipo de imágenes correspondientes a los bosques de robles, por las razones también expuestas en los mismos capítulos, se llega a la conclusión de que en este caso las características más apropiadas son las regiones, debido a que la estructura y composición de las imágenes así lo sugiere.

En ambos casos el interés se centra en la identificación de los troncos de los árboles, debido a que se trata del elemento diferenciador de la estructura de cualquier árbol.

Como bien es sabido, el mero hecho de analizar los píxeles o las regiones donde se supone que se ubican los troncos de los árboles, no es suficiente para llevar a cabo el subsiguiente proceso de correspondencia estereoscópica. Esto es así porque cada elemento de esa naturaleza requiere un análisis posterior, con el fin de conseguir su identificación de la mejor manera posible. De este modo, si cada característica se encuentra identificada en ambas imágenes por una serie de propiedades que le son inherentes, su futuro emparejamiento puede ser posible durante el proceso de correspondencia estereoscópica. Por tanto, dentro del proceso de segmentación se incluye la extracción de propiedades capaces de identificar y definir lo mejor posible las características extraídas, facilitando así el proceso posterior de correspondencia, tal y como se puso de manifiesto en relación al esquema de la figura 2.1.

En definitiva, el presente capítulo aborda la problemática de la segmentación, que incluye extracción de características y sus propiedades. La segmentación se aborda desde las dos perspectivas mencionadas, esto es, por un lado utilizando píxeles junto con sus descriptores, y por otro lado regiones, con su conjunto de descriptores asociados. En ambos casos, el proceso de segmentación se orienta de forma que el interés lo constituyen bien los píxeles de los troncos, en el primer caso, o bien las regiones asociadas a los troncos en el segundo. El resto de elementos presentes en la imagen, constituidos principalmente por otros tipos de texturas, se analizan mediante una serie de técnicas, cuyo objetivo se encamina hacia su identificación para su posterior exclusión, hasta donde sea posible, del proceso de correspondencia en visión estereoscópica.

Tanto las técnicas de identificación de las texturas para su exclusión como aquellas otras para la identificación de las texturas de interés se describen en el presente capítulo.

En la sección 3.2 se aborda el primer tipo de técnicas, basadas en píxeles, mientras que en la sección 3.3 se trata el problema de la identificación de texturas bajo la perspectiva de regiones.

## 3.2 Segmentación basada en píxeles

### 3.2.1 Generalidades

Como se ha mencionado reiteradamente, las imágenes analizadas bajo la perspectiva de la segmentación a nivel de píxel corresponden a imágenes procedentes de diferentes pinares. La figura 1.3 muestra una imagen representativa de este tipo de imágenes. Como se mencionó con anterioridad, el objetivo de este proceso de segmentación consiste en tratar de aislar los píxeles pertenecientes a los troncos de los árboles. No obstante, por las dificultades inherentes a este tipo de píxeles, según se explicó en el primer capítulo, más que extraer estos píxeles, de lo que realmente se trata es de eliminar el resto. Es decir, la idea consiste en eliminar aquellos píxeles que se pueden identificar como no pertenecientes a troncos. De este modo, el objetivo relativo al proceso de segmentación propuesto a continuación se centra precisamente en tratar de excluir los píxeles pertenecientes a uno de los tres tipos de texturas sin interés, que aparecen abundantemente en las imágenes, es decir: cielo, hierba en el suelo y hojas de los árboles. Los píxeles pertenecientes a dichas texturas son fácilmente identificables desde el punto de vista del observador humano, y relativamente más fáciles de extraer desde el punto de vista computacional, que los píxeles pertenecientes a las texturas que caracterizan los troncos.

Es evidente que la exclusión de las texturas carentes de interés en el análisis forestal resulta ciertamente útil, ya que los posibles errores que se pueden introducir durante el posterior proceso de correspondencia estereoscópica se reducen considerablemente. Este hecho justifica suficientemente la utilización de las técnicas descritas a continuación dentro del proceso de segmentación global para la clasificación e identificación de las mencionadas texturas de cara a su exclusión. No obstante, los procesos diseñados con tal propósito deben configurarse de forma que sean lo suficientemente conservadores para no excederse en la exclusión, evitando así la posibilidad de eliminar ciertas texturas pertenecientes a los propios troncos por su similitud con las texturas que se pretenden excluir.

Un hecho a tener en cuenta en este análisis es que las imágenes vienen caracterizadas por su representación en el modelo de color *RGB*, por las razones ya expuestas en los dos capítulos precedentes. Por otra parte, además de las observaciones

realizadas en dichos capítulos, una inspección más detallada de las texturas objeto de exclusión, permite extraer las siguientes consideraciones, que ayudan a enfocar el proceso de segmentación de la manera más apropiada posible:

- a) Aquellas áreas donde la densidad de hojas en los pinos es importante, muestran una alta variación de niveles de intensidad entre un píxel y los píxeles circundantes, en un entorno conocido técnicamente como *vecindad*. Por lo tanto, la utilización de métodos basados en la detección de texturas con este comportamiento resultan ser los más apropiados. Probablemente, uno de los más simples sea el procedimiento de naturaleza estadística basado en el cómputo de la varianza respecto de los niveles de intensidad de un píxel y los que le rodean. De este modo, considerando los niveles de intensidad en un área centrada en un píxel junto con la de sus vecinos, si dicha zona presenta un elevado valor de varianza, parece evidente que debe corresponderse con una zona caracterizada por un alto contenido de contraste y por tanto, probablemente se trata de una zona donde las hojas de los pinos se manifiestan abundantemente.
- b) Por el contrario, los niveles de intensidad correspondientes a la parte que se identifica como texturas procedentes del cielo, presentan una variabilidad de intensidad relativamente baja, lo que se traduce en unos valores de varianza bajos o muy bajos, tendiendo a cero. Se trata por tanto de áreas claramente homogéneas, donde un píxel se encuentra rodeado de otros píxeles con valores de intensidad prácticamente similares. Además, en este caso específico coinciden, como no podría ser de otro modo, con el hecho de que la componente dominante del espectro visible resulta ser la correspondiente a la banda azul. Siguiendo el razonamiento anterior, podría pensarse en utilizar la varianza como descriptor de este tipo de texturas, si bien asumiendo que en este caso se obtendrían valores bajos o nulos de varianza. Esto representa un serio inconveniente desde el punto de vista de nuestra problemática, debido a que en los propios troncos aparecen texturas con este tipo de comportamiento, es decir, zonas de intensidad homogénea donde la componente espectral azul resulta ser la dominante. Este es el caso de las texturas de los troncos que aparecen fuertemente iluminados por la incidencia directa de los rayos solares; en cuyo caso, esto nos llevaría a su exclusión del proceso de correspondencia, algo que a

todas luces no resulta deseable por tratarse de texturas pertenecientes a troncos y por tanto del máximo interés. Por ello, en el presente caso se hace necesario pensar en otro tipo de métodos orientados a la identificación de dichas texturas. Siguiendo esta línea de razonamiento y tal y como se ha expuesto en el capítulo dos, se ha optado por métodos de clasificación en el más amplio y clásico sentido del término.

- c) Con respecto a la hierba existente en el suelo, desde el punto de vista de su homogeneidad, su comportamiento es similar al caso de las texturas del cielo, aunque ciertamente en menor medida. En este caso, aparece una mayor variabilidad en cuanto a los niveles de intensidad se refiere, debido fundamentalmente a la aparición de zonas sombreadas frente a otras más iluminadas. Este nivel de variabilidad, aunque más elevado que en el caso del cielo, resulta claramente insuficiente de cara a su posible identificación por las mismas razones expresadas previamente respecto de la homogeneidad. Tanto en las zonas sombreadas como en las más iluminadas, en estas texturas la componente espectral dominante resulta ser la verde. Por tanto, una vez descartada la propiedad de homogeneidad, el análisis de la signatura espectral de este tipo de texturas se perfila como la opción más adecuada, ya que no existen texturas en los troncos con dominancia de la componente espectral verde, a diferencia de lo que ocurría con la azul.
- d) Sin más que observar las texturas procedentes de los troncos, se deduce claramente que son las más difíciles de identificar. Esto es así fundamentalmente por el hecho de recibir la iluminación bajo muy diversas condiciones. En efecto, debido a la posición del sol, la parte visible de determinados troncos reciben iluminación directa, lo que produce zonas fuertemente iluminadas, que llegan incluso a saturar el dispositivo de captura de la imagen generando niveles de intensidad elevados, los cuales se manifiestan con tonalidades tendentes al blanco. En este caso, como se ha indicado previamente, llega a existir un cierto dominio de la componente espectral azul sobre las otras dos en el modelo *RGB*. Este comportamiento puede fácilmente observarse en la mitad derecha de la imagen de la figura 1.3(a), donde uno de los troncos presenta este comportamiento en su totalidad, mientras que otros troncos lo manifiestan

parcialmente, debido principalmente a la aparición de zonas sombreadas e iluminadas alternantes, como consecuencia de las sombras producidas por las copas de los propios pinos. Por el contrario, cuando la luz incide de manera indirecta sobre la cara oculta de los troncos, la cara visible al sistema de medición aparece sombreada en la imagen como consecuencia de su iluminación difusa. Esto hace que las texturas de los troncos aparezcan con sus componentes espectrales totalmente diferentes a las anteriores. Este es el caso de los troncos situados en la parte izquierda de la imagen de la figura 1.3(a). En este caso los troncos no presentan alternancia de texturas iluminadas y sombreadas, ya que las copas de los árboles no producen alternancias de claro-oscuro como en el caso anterior.

Debido a la problemática anteriormente expuesta, cuya conclusión general resulta ser la inexistencia de una única textura descriptiva para los troncos, la identificación de los mismos a partir del análisis de las texturas se torna una ardua tarea desde el punto de vista computacional, que es el que nos concierne desde la perspectiva de esta investigación. La solución al problema pasa por la aplicación de técnicas apropiadas para abordar la problemática expuesta. En este sentido se propone la siguiente estrategia:

1. Identificación de las texturas de las hojas empleando técnicas de análisis de texturas basadas en medidas estadísticas. Estas técnicas pueden tratar con la alta variabilidad de la intensidad y se explican en la sección 3.2.2.
2. Métodos basados en aprendizaje, concretamente técnicas de clasificación para abordar la dominancia de las componentes espectrales verde y azul en la hierba y el cielo respectivamente. Se explican en la sección 3.2.3.
3. Las texturas identificadas con las estrategias anteriores se excluyen del proceso de correspondencia, quedando de este modo sólo las texturas supuestamente pertenecientes a los troncos, que son a las que se aplica el proceso de correspondencia a nivel de píxel, descrito en el capítulo cuatro.

La figura 3.1 representa una particularización del esquema general del proceso de visión estereoscópica mostrado en la figura 2.1, poniendo especial énfasis en la parte

de segmentación, que constituye el objetivo del presente capítulo y se sintetiza como sigue:

1. Sobre cada una de las imágenes involucradas en el proceso se aplican las técnicas de identificación de texturas para su exclusión.
2. De cada imagen se extraen los atributos de los píxeles resultantes que han de participar en el proceso de emparejamiento con los píxeles de la otra imagen del par estereoscópico, siendo naturalmente siempre el mismo conjunto de atributos en ambas imágenes.

Finalmente, la correspondencia o emparejamiento se realiza píxel a píxel entre ambas imágenes. Para ello se toma primeramente como referencia la imagen izquierda, de forma que para cada píxel bajo correspondencia en ella, se busca su homólogo en la imagen derecha. El mismo proceso se lleva a cabo en una segunda fase, tomando como referencia la imagen derecha y buscando los homólogos en la izquierda. En ambos casos, sólo participan, obviamente, los píxeles no excluidos durante el proceso de segmentación.

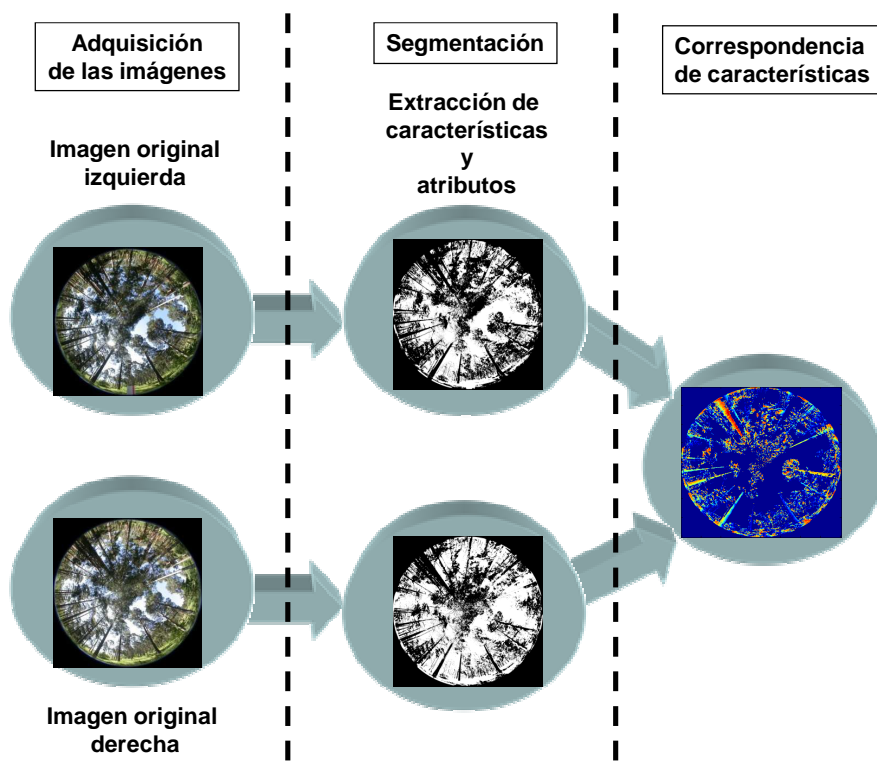


Figura 3.1 Esquema del proceso de segmentación y correspondencia basado en píxeles.



### 3.2.2 Identificación de texturas con alto contraste: Varianza

En la sección 2.2.2.2, al estudiar los métodos existentes en la literatura sobre la descripción de texturas se mencionaron una serie de ellos cuyo objetivo consistía en tratar de identificar ciertas estructuras repetitivas. Entre ellos se encontraban los siguientes: matrices de co-ocurrencia, filtros de Gabor, descomposición mediante *wavelets*, máscaras de Laws, coeficientes de energía, fractales o basados en la transformada del dominio (Fourier, transformada discreta del coseno). Lo cierto es que por las razones allí expuestas, se convino en que este tipo de descriptores no eran los apropiados para el tipo de texturas que aparecen en las imágenes objeto del presente estudio. En efecto, precisando un poco más al respecto, podemos añadir que las texturas representativas de las hojas de los árboles bajo estudio, no muestran en absoluto distribuciones de frecuencia ni patrones de textura repetitivos ni fácilmente identificables mediante los descriptores previos; antes bien, son áreas que presentan un alto contraste sin ninguna orientación espacial y con distribuciones totalmente aleatorias. Por tanto, la utilización de las técnicas anteriores, cuyo fundamento consiste en la descripción de la textura por la presencia de estructuras definidas, no resulta apropiada en nuestra problemática.

Se hizo pues, necesario orientar la investigación relativa a la identificación de estas texturas utilizando algún otro método. Afortunadamente, existen en la literatura varias técnicas empleadas hasta ahora para la identificación de texturas sin patrones predefinidos (Trias-Sanz y col., 2008). Su fundamento se basa en la comparación de ciertas magnitudes a partir de lo que se conoce como estadísticos de segundo orden (Gonzalez y Woods, 2008; Pajares y Cruz, 2007a). Estos métodos obtienen medidas de una serie de variables características de una determinada zona de textura en la imagen, tales como el grado de contraste, media, varianza, entropía, regularidad e incluso ciertas medidas de periodicidad, direccionalidad o aleatoriedad asociadas con la textura en un área restringida (Liu y Picard, 1996).

Tal y como ya se ha expuesto, dada la distribución aleatoria y el alto contraste de las texturas procedentes de las hojas de los pinos, los descriptores de texturas más apropiados son aquellos capaces de capturar la variación de alto contraste. En este sentido, la *varianza*, al tratarse de un descriptor del tipo momento estadístico de segundo orden, cumple ampliamente con el objetivo propuesto de captura del contraste.

Además, se trata de un descriptor relativamente sencillo y bien conocido, que ha sido utilizado en aplicaciones donde las medidas de contraste en imágenes constituyen un elemento esencial del proceso de análisis (Krotkov, 1987). También se han utilizado bajo la misma perspectiva de análisis del contraste, sistemas de visión estereoscópica con cámaras alineadas con sus ejes ópticos convergentes y con resultados ciertamente satisfactorios (Krotkov, 1989; Krotkov y col., 1990). En conclusión, para la identificación de las hojas de los pinos utilizamos la varianza estadística, que describimos a continuación, sobre un entorno de vecindad centrado en algún píxel característico de la textura objeto de identificación.

La aplicación del concepto de varianza, para el caso que nos ocupa, tiene su origen en las técnicas de identificación de texturas definidas en Gonzalez y Woods (2008) o Pajares y Cruz (2007a), que reproducimos aquí por su interés. Dado un píxel,  $i$  se considera una ventana centrada en  $i$  de tamaño  $n \times n$ , que define un entorno de vecindad alrededor de  $i$ . Por tanto, los píxeles en dicha ventana se sitúan sobre una región  $r$ -conectada espacialmente, donde  $r = n \times n - 1$ ; esto es, si por ejemplo  $n$  es 3, la vecindad es de  $3 \times 3$  y la región resulta ser 8-conectada. El objetivo, en cualquier caso, consiste en determinar si el píxel  $i$  pertenece o no a una zona de alto contraste con el propósito de catalogarlo como perteneciente a las texturas identificativas de las hojas de los pinos.

La aplicación práctica sobre la definición del concepto de alta variabilidad en la intensidad a través de la varianza se realiza como sigue. Sea  $z$  una variable aleatoria denotando niveles de intensidad en la región  $r$ -conectada y sea  $p(z_i)$ ,  $i = 0, 1, 2, \dots, L-1$ , el histograma correspondiente, donde  $L$  representa el número de niveles de intensidad distintos y posibles en la imagen. En imágenes como las analizadas en este trabajo de investigación, con niveles de intensidad representados por 8 bits en cada canal espectral,  $L$  toma el valor de 256. La varianza se define como:

$$\sigma^2(z) = \sum_{i=0}^{L-1} (z_i - \bar{m})^2 p(z_i) \quad (3.1)$$

donde  $\bar{m}$  es el valor de intensidad medio de  $z$ , es decir, el nivel de intensidad medio calculado mediante la ecuación (3.2).

$$\bar{m} = \sum_{i=0}^{L-1} z_i p(z_i) \quad (3.2)$$

De (3.1) y según Gonzalez y Woods (2008) o Pajares y Cruz (2007a), un coeficiente capaz de medir el contraste de intensidad en la región  $r$ -conectada, normalizado en el rango  $[0,+1)$  se define como:

$$R = 1 - \frac{1}{1 + \sigma^2(z)} \quad (3.3)$$

A partir de la ecuación anterior, puede fácilmente deducirse que  $R$  es cero para áreas de intensidad constante, donde la varianza es nula, y se aproxima a +1 para valores grandes de  $\sigma^2(z)$ , es decir, áreas de alto contraste. Como se ha mencionado en los capítulos precedentes y se detalla en el capítulo cinco, las imágenes originales utilizadas en este trabajo se representan en el espacio de color  $RGB$ ; si bien para la medida del contraste a través del coeficiente  $R$  se realiza la correspondiente transformación al espacio de color  $HSI$  (*Hue, Saturation, Intensity*) (Gonzalez y Woods, 2008). En este espacio, la imagen de intensidad  $I$  es la que se usa para calcular la varianza y los valores de intensidad  $z_i$  oscilando en el rango  $[0,255]$  con  $L = 256$ . Bien es cierto que podrían haberse utilizado las tres componentes espectrales del modelo  $RGB$ , e identificado zonas de alto contraste en cada una de las componentes espectrales  $R$ ,  $G$  y  $B$ ; no obstante, diversos experimentos llevados a cabo en esta línea nos han permitido concluir que los resultados así obtenidos no mejoran sustancialmente los conseguidos con  $I$  a través del modelo  $HSI$ , e incluso en algunos casos empeoran. Esto se debe principalmente a que el propio modelo  $HSI$  engloba en su componente de intensidad la información procedente de los tres canales espectrales  $R$ ,  $G$ ,  $B$ ; por ello, dado que bajo el modelo  $HSI$  sólo se procesa la componente  $I$ , frente a las tres que requiere el modelo  $RGB$ , se ha preferido optar por el primero al producir resultados similares y un mejor rendimiento desde el punto de vista computacional, aún a pesar de que la propia transformación a  $HSI$  requiere un determinado coste, que en ningún momento es mayor que el que resulta de procesar las tres componentes  $R$ ,  $G$ ,  $B$  por separado.

Finalmente, el criterio para identificar el píxel  $i$  como perteneciente a un área con alto nivel de contraste, y por tanto su pertenencia a la textura característica de las

zonas de hojas, se establece bajo el criterio de que el coeficiente  $R$  debe tomar un valor mayor que un determinado umbral  $T_1$ , fijado a 0.8 en nuestro trabajo mediante diferentes estudios experimentales basados en pruebas de ensayo y error. No obstante, el criterio adoptado para fijar dicho valor umbral (relativamente elevado si tenemos en cuenta que el valor máximo es +1) se basa en que cabe la posibilidad de aparición de ciertas zonas con un valor de  $R$  alto que no pertenezcan a las texturas características e identificativas de las hojas, y sí a zonas propias de los troncos de los árboles. Esto puede ocurrir con cierta facilidad en el caso de las imágenes bajo estudio procedentes de entornos no estructurados de exterior. En efecto, debido a las condiciones ambientales bajo las que se capturan dichas imágenes, principalmente la iluminación, pueden existir con relativa facilidad zonas de texturas en los troncos donde el contraste de intensidad sea elevado. Por tanto, con un valor de umbral bajo, podrían excluirse del posterior proceso de correspondencia estereoscópica una serie de píxeles pertenecientes a troncos, lo que conduciría a una situación no deseada.

### 3.2.3 Identificación de texturas homogéneas: Clasificación

#### 3.2.3.1 Planteamiento del enfoque

Como se ha mencionado previamente, en los estudios realizados con el tipo de imágenes procedentes de los pinos, además de las texturas con elevado contraste, descritas en la sección anterior, existen en las imágenes otros dos tipos de texturas relevantes a identificar por el interés suscitado de cara a su exclusión del proceso de correspondencia posterior, a saber: hierba del suelo y cielo. Su relativa homogeneidad y por consiguiente bajo contraste en cuanto a los niveles de intensidad se refiere, no permite la aplicación segura del criterio definido previamente, ya que en este caso su aplicación consistiría en determinar qué píxeles poseen valores de  $R$  bajos por aplicación de la ecuación (3.3) y por tanto por debajo del umbral anterior  $T_1$ . Se da la circunstancia, de que en los troncos muchas de las texturas que aparecen cumplen este criterio. Por tanto, utilizado con este propósito, es evidente que estas texturas pertenecientes a los troncos, quedarían automáticamente excluidas del proceso de correspondencia, algo que a todas luces nos conduce a una situación indeseable, ya que los troncos de los árboles constituyen la parte esencial de la correspondencia.

Por lo tanto, para la identificación de las mencionadas dos texturas, recurrimos a técnicas clásicas de clasificación. Dado que el objetivo de este trabajo no consiste en la investigación en este campo, se eligen dos métodos que han sido probados satisfactoriamente en la clasificación de texturas de imágenes naturales en su versión combinada (Guijarro, 2009; Guijarro y col., 2008, 2009; Guijarro y Pajares, 2009; Pajares y col., 2009, 2010). Si bien es cierto que las imágenes sobre las que se han aplicado proceden de sensores multiespectrales aerotransportados, poseen no obstante dos características comunes con las imágenes analizadas en este trabajo: *a)* las bandas espectrales analizadas son las que corresponden al espectro visible, como en el caso que nos ocupa según el modelo de color *RGB* y *b)* proceden de entornos naturales de exterior y por tanto no estructurados. En consecuencia, presentan muchos problemas similares a los que nos encontramos en el caso de las imágenes forestales, principalmente el que se deriva de las diferencias de iluminación y por consiguiente de los diferentes cambios en los niveles de intensidad de los píxeles dentro de una misma textura. Los dos métodos de clasificación anteriormente mencionados y utilizados como referencia en la investigación llevada a cabo en este trabajo son, por un lado el método conocido como Agrupamiento Borroso o en terminología inglesa *Fuzzy Clustering (FC)* y por otro el clásico estimador paramétrico Bayesiano (*PB*). Resulta bien conocido, a partir de la investigación plasmada y constatada en los trabajos anteriormente citados, que la tendencia actual en aplicaciones de clasificación se orienta en general hacia la combinación de clasificadores, algo que ya fue apuntado en algunos trabajos precursores clásicos tales como en Yager (1988) o más recientemente en Kuncheva (2004). Por estas razones, la propuesta que se formula en el presente trabajo se fundamenta en lo anterior y se concreta en la combinación de los dos clasificadores individuales *FC* y *PB* aplicada al reconocimiento e identificación de texturas. No obstante, se ha optado por una estrategia combinada simple, verificando que no es necesario recurrir a métodos más complejos de combinación como los descritos en cualquiera de las referencias citadas previamente. Esta opción se ha elegido teniendo en cuenta que las dos texturas a identificar para su exclusión no presentan grandes dificultades y que en cualquier caso, aunque existan píxeles con cierto grado de dificultad en su identificación, esto no entraña un grave problema, pues el posterior proceso de correspondencia se encarga de emparejar o no los mencionados píxeles, con lo que el problema quedaría paliado en parte.

Tanto el método de clasificación *FC* como *PB* son de naturaleza supervisada, por lo que requieren de una fase de *entrenamiento* previa a la de *decisión*. A continuación se detallan brevemente los fundamentos teóricos relativos a ambas fases.

En la propuesta formulada en este trabajo, durante la fase de entrenamiento, también llamada de aprendizaje, se estiman y se aprenden una serie de parámetros propios de ambos clasificadores, siendo posteriormente en la fase de decisión donde realmente se lleva a cabo la combinación. Los resultados de la segmentación obtenidos sobre las imágenes de prueba durante dicha fase de decisión, tanto cuando se utilizan ambos clasificadores por separado como cuando se aplican combinados, se proporcionan en el capítulo cinco. No obstante, queda todavía un amplio margen de investigación de futuro, particularmente en lo que se refiere al análisis de nuevos tipos de texturas. Este es el caso de las ideas apuntadas en el apéndice B donde queda patente la necesidad de desarrollar nuevas estrategias de clasificación para determinados tipos de entornos forestales como el que allí se describe. Esta línea de investigación surge como consecuencia del intento de ampliación de los métodos a nuevos entornos forestales, intuyendo, lógicamente, que no será la única cuando se analicen otros tipos de imágenes.

A continuación se describen los procesos involucrados en sendas fases de aprendizaje y decisión.

### 3.2.3.2 Fase de entrenamiento o aprendizaje

#### a) Agrupamiento Borroso (*FC*)

En esta fase se comienza con la observación de un conjunto  $X$  de  $n$  patrones de entrenamiento, es decir,  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathfrak{R}^d$ . Cada muestra se asigna a una clase dada, que identificamos como  $c_j$ , donde el número de clases posibles es  $c$ , y por consiguiente,  $j = 1, 2, \dots, c$ . En este trabajo el número de clases objeto de estudio es dos, por tanto,  $c = 2$ . Por simplicidad, se identifica la clase  $c_1$  con el cielo y la clase  $c_2$  con la hierba del suelo. Los patrones  $\mathbf{x}_i$  pertenecientes al conjunto  $X$  representan píxeles en el espacio de color *RGB*. Sus componentes son los correspondientes valores espectrales  $R$ ,  $G$ ,  $B$ . Esto significa que la dimensión del espacio  $\mathfrak{R}$  es tres y en consecuencia,  $d = 3$ .

El proceso de aprendizaje recibe los patrones de entrenamiento como entrada, que han sido previamente asignados como pertenecientes a una de las dos clases objeto de estudio,  $c_1$  y  $c_2$ . Según Zimmermann (1991) o Duda y col. (2000), el clasificador  $FC$  calcula para cada  $x_i$  en la iteración  $k$  su grado de pertenencia  $\mu_{ij}$  a una determinada clase  $j$  y actualiza los centros de la clase,  $v_j \in \mathfrak{R}^d$ , de acuerdo con las muestras asignadas a esa clase según las ecuaciones (3.4) y (3.5).

$$\mu_{ij}(k+1) = \frac{1}{\sum_{r=1}^c (d_{ij}(k)/d_{ir}(k))^{2/(e-1)}} \quad (3.4)$$

$$v_j(k+1) = \frac{\sum_{i=1}^n \mu_{ij}^e(k) x_i}{\sum_{i=1}^n \mu_{ij}^e(k)} \quad (3.5)$$

$d_{ij}^2 \equiv d^2(x_i, v_j)$  es la distancia Euclídea al cuadrado. El exponente  $e$  recibe el nombre de peso exponencial (Zimmermann, 1991; Bezdek, 1981). Al tratarse de un proceso iterativo, es necesario establecer el criterio de parada del mismo, que se establece en nuestro caso, según se cumpla una cualquiera de las dos condiciones siguientes:

- 1) Los grados de pertenencia entre dos iteraciones consecutivas no varía en más de una cierta cantidad  $\varepsilon$ . La variación se determina según la siguiente relación  $\|\mu_{ij}(k+1) - \mu_{ij}(k)\| < \varepsilon \quad \forall ij$ . Esto también es equivalente a establecer que los centros no varían entre iteraciones consecutivas. En ambos casos, bien sea porque los centros o los grados de pertenencia no varían de una iteración a la siguiente, se dice que el proceso ha logrado la convergencia.
- 2) Se alcanza un número de iteraciones establecido de antemano  $k_{max}$ . En este caso no se puede asegurar la llegada a la convergencia, por lo que el proceso de aprendizaje podría ser, en este supuesto, mejorable. Dicho número máximo de iteraciones se establece para evitar que ante la falta de convergencia el proceso no finalice.

En nuestros experimentos, los valores establecidos mediante ensayo y error han sido:  $e = 2.1$ ,  $\varepsilon = 0.01$  y  $k_{max} = 50$ , que han resultado ser suficientemente satisfactorios, tal y como se describe en la sección 5.2.1.

En cualquier caso, el método iterativo requiere la inicialización bien de los centros de cada clase para comenzar con la ecuación (3.4) o bien de los grados de pertenencia, en cuyo caso el proceso iterativo comenzaría en la ecuación (3.5). En cualquiera de los dos casos tanto los grados de pertenencia como los centros de cada clase se actualizan alternativamente a través de las mencionadas ecuaciones (3.4) y (3.5). En el presente trabajo se ha elegido la opción de comenzar el proceso mediante la inicialización de los centros de las clases a través del procedimiento pseudoaleatorio descrito en Balasko y col. (2008) empezando en la iteración  $k = 1$ . Dicho procedimiento se concreta en los tres puntos siguientes:

- 1) Se realiza una transformación lineal  $Y = f(X)$  de las muestras de entrenamiento para que los valores se sitúen en el intervalo cerrado  $[0,1]$ .
- 2) Se inicializan los centros como sigue:  $\mathbf{v} = 2D\overline{\mathbf{M}} \circ \mathbf{R} + D\overline{\mathbf{m}}$ , donde  $\overline{\mathbf{m}}$  es el vector media para los valores de las muestras de entrenamiento transformadas en  $Y$  y  $\overline{\mathbf{M}} = \max(\text{abs}(Y - \overline{\mathbf{m}}))$ , ambos de tamaño  $1 \times d$ ;  $D = [1 \dots 1]^T$  con tamaño  $c \times 1$ ;  $\mathbf{R}$  es una matriz  $c \times d$  de números aleatorios en  $[0,1]$ ; la operación  $\circ$  denota la multiplicación elemento por elemento y  $T$  se refiere en este caso a la transpuesta.
- 3) Una vez finalizado el proceso de entrenamiento, por cualquiera de los criterios de parada definidos previamente, se aplica la transformada inversa a la realizada en el punto primero, para devolver los valores de los centros de clase al rango original.

#### b) Estimador paramétrico Bayesiano (PB)

Este estimador, basado en la teoría de la probabilidad de Bayes, asume una distribución de densidad de probabilidad conocida, generalmente Gaussiana, para cada clase expresada como sigue (Duda y col. 2000; Pajares y Cruz, 2007a,b):

$$p(\mathbf{x} | c_j) = \frac{1}{(2\pi)^{d/2} |C_j|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{v}_j)^T C_j^{-1} (\mathbf{x} - \mathbf{v}_j) \right] \quad (3.6)$$

donde los parámetros a ser estimados son la media  $\mathbf{v}_j$  y la matriz de covarianza  $C_j$ , ambas para cada clase  $c_j$ . La estimación de dichos parámetros se establece mediante el



conocido proceso de máxima verosimilitud, concluyendo finalmente que la mejor estima para ambos viene dada por la ecuación siguiente:

$$\mathbf{v}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} \mathbf{x}_k \quad C_j = \frac{1}{n_j - 1} \sum_{k=1}^{n_j} (\mathbf{x}_k - \mathbf{v}_j)(\mathbf{x}_k - \mathbf{v}_j)^T \quad (3.7)$$

donde  $T$  denota la transpuesta y  $n_j$  es el número de muestras pertenecientes a la clase  $c_j$ . Obsérvese cómo para los centros la mejor estima coincide exactamente con la media de los valores de la clase.

### 3.2.3.3 Fase de decisión

#### a) Agrupamiento Borroso (FC)

Tras la fase de entrenamiento, se dispone de los centros de las clases, que han sido convenientemente aprendidos mediante el proceso de entrenamiento previo. En la fase de decisión el objetivo consiste en clasificar una nueva muestra  $\mathbf{x}_s \in \mathcal{R}^d$ , cuya pertenencia a las clases  $c_j$  resulta desconocida.

Obviamente, las nuevas muestras, al igual que en la fase de entrenamiento, representan píxeles en la imagen. Los vectores representativos de estas muestras se caracterizan por ser vectores tridimensionales, cuyas componentes son las componentes espectrales  $R, G, B$ , que definen el píxel que representa.

El clasificador  $FC$  realiza el proceso de decisión como sigue. Dada la muestra  $\mathbf{x}_s$  se obtienen para la misma los grados de pertenencia a cada clase según la ecuación (3.4), teniendo presente que los centros de las clases son los aprendidos previamente. La muestra, y por tanto el píxel que representa, se clasifica según la siguiente regla:  $\mathbf{x}_s \in c_j$  si  $\mu_{sj} > \mu_{sh}$  y  $\mu_{sj} > T_2$  de lo contrario el píxel permanece sin clasificar. Es evidente que según esta regla, se exige que el grado de pertenencia de  $\mathbf{x}_s$  a la clase  $c_j$  sea el mayor de todos ellos y además que dicho grado supere un determinado valor umbral  $T_2$ . La primera premisa es obligada según el criterio generalmente aplicado con este clasificador. La segunda se introduce como consecuencia de nuestro problema en particular. Con ella lo que se pretende es clasificar sólo aquellos píxeles que ofrezcan un elevado grado de pertenencia, dejando el resto sin clasificar para el subsiguiente proceso de correspondencia estereoscópica. Entre éstos se encontrarán los píxeles pertenecientes

a los troncos, que constituyen el objeto de interés. De esta forma, se pretende clasificar sólo los píxeles sobre los que no haya duda respecto de su pertenencia a una de las dos clases que constituyen el objetivo de su exclusión del posterior proceso de correspondencia. El umbral  $T_2$  se ha fijado al valor 0.8 tras diversos experimentos de ensayo y error. Como puede apreciarse, se trata de un valor relativamente alto, que sólo identifica píxeles con un alto grado de pertenencia a cualquiera de las dos clases  $c_1$  o  $c_2$ , garantizando así que el píxel clasificado pueda excluirse con cierto grado de confianza.

La figura 5.7(a) y (b) muestra el resultado de aplicar a la imagen de la figura 1.3(a) y (b) respectivamente el proceso de segmentación descrito hasta este momento, es decir, extrayendo las texturas correspondientes a las hojas de los pinos mediante el proceso planteado en la sección 3.2.2 basado en la varianza, y el proceso de decisión basado en el clasificador  $FC$ , descrito en esta misma sección. Las zonas etiquetadas en blanco son los píxeles identificados como pertenecientes a uno de los tres tipos de texturas que quedan fuera del objeto de interés, es decir, hojas, cielo y la hierba sobre el suelo. Por el contrario, las zonas de negro dentro del círculo de definición de la imagen, son los píxeles que pasarán al proceso de correspondencia estereoscópica. Entre estos últimos, obviamente, se encuentran los píxeles pertenecientes a los troncos, que constituyen nuestro objeto de interés. Como puede observarse, la mayoría de los troncos aparecen en color negro. Existen troncos, etiquetados como blancos, que representan los árboles situados muy lejos del sensor, concretamente fuera del área de los 25m de distancia, que se considera como la distancia umbral para el análisis. Dichos troncos no se consideran por situarse fuera de nuestro interés, tal y como se detalla en el capítulo cinco.

#### b) Estimador paramétrico (PB)

La fase de decisión correspondiente a este clasificador se basa en una regla similar a la utilizada en el caso del  $FC$ . Así, dada la muestra  $\mathbf{x}_s$ , se obtienen para la misma las probabilidades de pertenencia a cada clase según la ecuación (3.6), teniendo presente que los centros y las matrices de covarianza asociadas a dichas clases son los parámetros aprendidos previamente. La muestra, y por tanto el píxel que representa, se clasifica según la siguiente regla:  $\mathbf{x}_s \in c_j$  si  $p(\mathbf{x}_s | c_j) > p(\mathbf{x}_s | c_h)$  y  $p(\mathbf{x}_s | c_j) > T_2$  de lo contrario el píxel permanece sin clasificar. Sobre el umbral  $T_2$  se aplica el mismo razonamiento que en el caso anterior relativo al clasificador  $FC$ .

La figura 5.8(a) y (b) muestra el resultado de aplicar a la imagen de la figura 1.3(a) y (b) respectivamente el proceso de segmentación descrito hasta este momento, es decir, extrayendo las texturas correspondientes a las hojas de los pinos mediante el proceso descrito en la sección 3.2.2 a través de la varianza estadística y el proceso de decisión basado en el clasificador *PB*, descrito en esta misma sección. En relación a la imagen 5.7, se aprecia un aumento del número de píxeles pertenecientes a las texturas de cielo y suelo, cuya discusión se pospone al capítulo cinco.

### c) Combinación de Clasificadores

Como se ha mencionado reiteradamente, el objetivo de la combinación de clasificadores consiste en aprovechar y unir las ventajas individuales de cada uno de los clasificadores que participan en la combinación durante la fase de decisión, haciendo bueno el dicho popular de que la unión hace la fuerza.

Durante la combinación el clasificador *FC* calcula los grados de pertenencia de  $x_s$  a cada clase según la ecuación (3.4), mientras que *PB* calcula las probabilidades de que  $x_s$  pertenezca a una de las dos clases según la ecuación (3.6). Ambos, probabilidades y grados de pertenencia, constituyen las salidas de los clasificadores individuales relativas al proceso de decisión cuyos valores varían en el rango [0,1]. Existen numerosas formas para combinar dichas salidas, una de las más simples, aunque no por ello menos eficaz, es la *regla de la media* (Kuncheva, 2004). Tax y col. (2000) realizan un estudio comparativo sobre el comportamiento de clasificadores combinados utilizando como regla de combinación la *media* y la *multiplicación*. Los resultados obtenidos en dicho trabajo permiten concluir que el mejor comportamiento se obtiene con el producto cuando las muestras pertenecen a espacios de características independientes, mientras que en espacios de características completamente dependientes el rendimiento es el mismo tanto para el producto como la regla de la media.

En el caso que nos ocupa, nuestras características, es decir los píxeles, se ubican en el espacio de color *RGB*, que presenta una alta correlación entre las componentes espectrales *R*, *G* y *B*, tal y como se describe en los estudios llevados a cabo por Littmann y Ritter (1997) o Cheng y col. (2001). Una alta correlación implica que si la intensidad varía, las tres componentes espectrales cambiarán en consecuencia. Por tanto, debido a esta correlación existente en nuestro espacio de color, se elige la media como

la regla más apropiada para la combinación, que se sintetiza como sigue:  $m_{sj} = 0.5(\mu_{sj} + p(\mathbf{x}_s | c_j))$ . Ahora, el píxel representado por  $\mathbf{x}_s$  se clasifica como perteneciente a una de las dos clases según la regla de decisión siguiente:  $\mathbf{x}_s \in c_j$  si  $m_{sj} > m_{sh}$  y  $m_{sj} > T_2$  de lo contrario el píxel permanece sin clasificar. El umbral  $T_2$  se introduce con la misma finalidad que en el caso de los clasificadores individuales, esto es, para garantizar el hecho de que  $\mathbf{x}_s$  representa a un píxel con amplias garantías de pertenecer a una de las clases objeto de exclusión durante el proceso de correspondencia.

La figura 5.9(a) y (b) muestra el resultado de aplicar el proceso de segmentación a la imagen de la figura 1.3(a) y (b) respectivamente mediante la combinación de los clasificadores individuales.

### 3.2.4 Extracción de atributos

Siguiendo el esquema de la figura 2.1, el proceso de segmentación, tras la extracción de características continúa con el etiquetado y extracción de atributos procedentes de las características. El proceso de segmentación se aplica paralelamente sobre las dos imágenes del par estereoscópico bajo procesamiento.

En el caso que nos ocupa, el etiquetado se refiere al hecho de identificar todos aquellos píxeles que van a formar parte del proceso de correspondencia estereoscópica. En el caso de las figuras 5.7 a 5.9, este proceso se asocia con el marcado de los píxeles con el valor lógico “0”, esto es, negro.

Pues bien, para cada uno de los píxeles marcados con la etiqueta cero, tanto en la imagen izquierda como derecha del par estereoscópico, se obtienen los atributos que se relacionan a continuación: a) correlación; b) color; c) textura; d) módulo del gradiente; e) ángulo del gradiente y f) Laplaciana. Todos ellos tienen como finalidad obtener el máximo grado de similitud entre los píxeles bajo correspondencia de cara al emparejamiento entre píxeles de las imágenes izquierda y derecha del par estereoscópico. A continuación se describe brevemente cada uno de ellos.

### 3.2.4.1 Correlación

La técnica de correlación probablemente sea una de las más utilizadas en sistemas de visión estereoscópica convencional para sistemas de navegación autónoma de robots en entornos no estructurados de exterior, tal es el caso de la aplicación desarrollada por Rovira-Más y col. (2008) para campos de cultivo o Matthies y col., (2007) en sistemas de exploración planetaria, entre otras muchas. Nuestros entornos forestales se catalogan también como no estructurados y de exterior, por lo que el coeficiente de correlación se presenta como un atributo de interés.

Dado un píxel en la imagen izquierda, se establece un entorno de vecindad definiendo una ventana de dimensión  $M \times N$  alrededor del mismo. En cada píxel candidato en la imagen derecha se abre una ventana con las mismas dimensiones, de forma que entre ambas ventanas se calcula el coeficiente de correlación como sigue (Barnea y Silverman, 1972; Koschan y Abidi, 2008; Klaus y col., 2006):

$$C = \frac{\sigma_{ID}^2}{\sqrt{\sigma_I^2 \sigma_D^2}} \quad (3.8)$$

donde los subíndices  $I$  y  $D$  se refieren a las imágenes izquierda y derecha, respectivamente,  $\sigma_I^2$  y  $\sigma_D^2$  representan la varianza de los niveles de intensidad en las correspondientes ventanas, y  $\sigma_{ID}^2$  es la covarianza de los niveles de intensidad entre las ventanas izquierda y derecha. Estos coeficientes están definidos en las siguientes ecuaciones:

$$\sigma_k^2 = \sum_{i=1}^M \sum_{j=1}^N \frac{(I_k(i, j) - \mu_k)^2}{MN}; \quad k = I, D \quad (3.9)$$

$$\sigma_{ID}^2 = \sum_{i=1}^M \sum_{j=1}^N \frac{(I_I(i, j) - \mu_I)(I_D(i, j) - \mu_D)}{MN} \quad (3.10)$$

En las anteriores expresiones la intensidad en cada píxel  $(i, j)$  viene dada por  $I_k(i, j)$  y la media del nivel de intensidad en el entorno de vecindad o ventana resulta ser  $\mu_k$ . En ambos casos, el subíndice  $k$  indica si se trata de la imagen izquierda o derecha. Dado que en nuestro caso las imágenes originales se encuentran en el modelo de color  $RGB$ ,

el coeficiente de correlación se obtiene sobre la imagen de intensidad perteneciente al modelo de color *HSI*, previa transformación de la imagen a este modelo desde el *RGB*.

### 3.2.4.2 Color

Un píxel cualquiera  $i$  en la imagen izquierda, localizado en una determinada posición espacial  $(x,y)$  en la imagen, se define por su vector de características  $\mathbf{x}_i$ , cuyas componentes son los valores de las correspondientes tres bandas espectrales en el modelo de color *RGB*, esto es,  $\mathbf{x}_i = (R_i, G_i, B_i)$ ; para este píxel en la imagen derecha existen una serie de candidatos situados en sus correspondientes localizaciones espaciales, donde cada uno de ellos, identificados genéricamente como  $j$ , tiene asociado su correspondiente vector de características  $\mathbf{x}_j = (R_j, G_j, B_j)$ ; en este caso,  $\mathbf{x}_i$  y  $\mathbf{x}_j$  son los atributos descriptivos de los píxeles en ambas imágenes. El color es un potente atributo que se ha utilizado en diferentes trabajos relativos al establecimiento de similitudes entre píxeles (Koschan y Abidi; 2008) y más específicamente en correspondencia estereoscópica, cual es el caso de Klaus y col. (2006) que lo han utilizado en combinación con el módulo del gradiente. En este último trabajo, la medida de similitud entre dos píxeles con vectores  $\mathbf{x}_i$  y  $\mathbf{x}_j$  es la *distancia de Manhattan* definida como

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{X=R,G,B} |X_i - X_j|.$$

### 3.2.4.3 Textura

En este caso, el atributo de textura asociado a un píxel se obtiene mediante la aplicación de la ecuación (3.3), obteniéndose así un coeficiente de textura para cada píxel, tanto en la imagen izquierda como en la imagen derecha bajo emparejamiento. Dados dos píxeles  $i$  y  $j$  pertenecientes a las imágenes izquierda y derecha respectivamente, se obtienen sus correspondientes coeficientes de textura, designados ahora como  $R_i$  y  $R_j$  según dicha ecuación. Los valores de estos coeficientes se calculan sobre la imagen de intensidad del modelo *HSI*, obteniendo una medida de similitud dada por el valor absoluto de la diferencia entre ambos coeficientes.

### 3.2.4.4 Módulo y dirección del gradiente

Lew y col. (1994) han utilizado satisfactoriamente tanto el módulo del gradiente como su dirección como propiedades o atributos de los píxeles. Klaus y col. (2006) también lo han utilizado en combinación con el atributo de color.

Su cómputo resulta relativamente sencillo a partir del concepto *primera derivada* (Pajares y Cruz, 2007a). En efecto, el gradiente de una imagen  $f(x,y)$  en un píxel localizado en la posición espacial  $(x,y)$  se define como un vector bidimensional dado por la ecuación (3.11).

$$\mathbf{G}[f(x, y)] = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} f(x, y) \\ \frac{\partial}{\partial y} f(x, y) \end{bmatrix} \quad (3.11)$$

donde el vector  $\mathbf{G}$  apunta en la dirección de variación máxima de  $f$  en el píxel  $(x,y)$  por unidad de distancia con la magnitud y dirección dadas por:

$$|\mathbf{G}| = \sqrt{G_x^2 + G_y^2}; \quad \phi(x,y) = \tan^{-1} \frac{G_y}{G_x} \quad (3.12)$$

La obtención de las componentes del gradiente puede realizarse de muy diferentes maneras. En este trabajo de investigación se materializa mediante la aplicación de los conocidos operadores de Sobel:

$$G_x \equiv \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad G_y \equiv \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (3.13)$$

El módulo como la dirección son atributos que se obtienen tanto en el píxel de la imagen izquierda,  $i$ , como en los de la derecha,  $j$ , bajo el proceso de emparejamiento y aplicándose los operadores sobre la imagen de intensidad tras su transformación desde el plano  $RGB$  al  $HSI$ . Como en los casos anteriores, una medida de similitud entre  $i$  y  $j$  puede establecerse sencillamente calculando la diferencia en valor absoluto de los valores del módulo y dirección del gradiente por separado. Esto es, obteniendo una medida de similitud para el módulo y otra para la dirección.

### 3.2.4.5 Laplaciana

Se trata de otro atributo utilizado en Lew y col. (1994) para correspondencia en visión estereoscópica. La Laplaciana de una función 2-D  $f(x,y)$  es un operador segunda derivada definido como:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (3.14)$$

Al igual que en el caso del gradiente, la ecuación (3.14) se puede implementar en forma digital de varias formas, siendo la proporcionada en la ecuación (3.15) la de uso más generalizado por su buen comportamiento, ya que se trata de un operador isótropo, lo que significa que captura las propiedades en todas las direcciones del espacio, siendo ésta la razón de su utilización en este trabajo:

$$\nabla^2 f \equiv \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (3.15)$$

Al igual que en el caso anterior, se aplica sobre la imagen de intensidad tras su transformación al modelo de color *HSI*. Una medida de similitud entre los valores de la Laplaciana para dos píxeles  $i$  y  $j$  pertenecientes a las imágenes izquierda y derecha respectivamente, se puede establecer simplemente mediante el valor absoluto de la diferencia, al igual que para el resto de atributos descritos previamente.

### 3.2.5 Resumen del proceso de segmentación basado en píxeles

En síntesis, el proceso global sobre segmentación basado en píxeles se puede resumir como sigue:

- 1) En sendas imágenes izquierda y derecha se extraen las texturas con alto contraste, hojas de los pinos, mediante la aplicación de la ecuación (3.3).
- 2) También sobre ambas imágenes, se extraen las texturas con un cierto grado de homogeneidad, que serán excluidas del proceso de correspondencia, hierba del suelo y cielo. Para ello se aplica uno de los procedimientos de clasificación simple (*FC*, *PB*) o el combinado, descritos en la sección 3.2.3.



Como se ha mencionado previamente, en la investigación realizada en este trabajo se ha utilizado éste último.

- 3) Para el proceso de correspondencia posterior, se utilizan como atributos los seis descritos en la sección 3.2.4.

A modo de síntesis, en la tabla 3.1 se muestran los valores más significativos de los umbrales y parámetros involucrados en el proceso de segmentación basado en píxeles, distinguiendo si son relativos a la fase de entrenamiento o a la de decisión.

Tabla 3.1 Parámetros y umbrales implicados en el proceso de segmentación basado en píxeles

Fase	Parámetro/Umbral	Descripción
Aprendizaje	$e = 2.1$ $\varepsilon = 0.01$ $k_{max} = 50$	Parámetros involucrados en las ecuaciones (3.4) y (3.5) en el proceso de aprendizaje del clasificador <i>FC</i>
Decisión	$T_1 = 0.8$	Umbral para decidir si una determinada textura pertenece a zona de hojas o no, según la sección 3.2.2
	$T_2 = 0.8$	Umbral para decidir si una determinada textura pertenece a las clases suelo o cielo, según las reglas de decisión de los clasificadores <i>FC</i> y <i>PB</i> , sección 3.2.3.3

### 3.3 Segmentación basada en regiones

#### 3.3.1 Generalidades

Como se mencionó con anterioridad, el objetivo fundamental de la segmentación consiste en extraer automáticamente las regiones asociadas a los troncos y sus propiedades, de modo que estas regiones puedan corresponderse en ambas imágenes.

A diferencia de la segmentación basada en píxeles donde las características son los píxeles que no han sido clasificados como pertenecientes a las texturas objeto de exclusión, ahora las características de interés son las regiones que sí se van a tratar de identificar durante el proceso de segmentación. El objetivo primordial ahora consiste en asociar regiones a troncos de árboles, de forma que a partir de esas regiones puedan obtenerse las propiedades de los mismos como paso previo hacia la correspondencia, en este caso por regiones.

La figura 3.2 representa una particularización del esquema general del proceso de visión estereoscópica mostrado en la figura 2.1, poniendo especial énfasis en la parte de segmentación, que constituye, como en el caso de los píxeles, el objetivo del presente capítulo. Podemos sintetizarlo como sigue:

- 1) En primer lugar se realiza un proceso de extracción de regiones, que se aplica sobre ambas imágenes. De este modo, para cada imagen se obtiene un conjunto de regiones, cada una de las cuales se asocia con un tronco.
- 2) Para cada región en cada imagen se aplican técnicas de etiquetado, obteniéndose una serie de propiedades descriptivas de las mismas, que se utilizarán posteriormente en el proceso de correspondencia entre regiones. La decisión sobre el emparejamiento entre dos regiones se realiza por similitud de propiedades entre una región de la imagen izquierda y el conjunto de regiones candidatas en la imagen derecha, y viceversa, entre una región en la imagen derecha y el conjunto de regiones candidatas en la imagen izquierda.

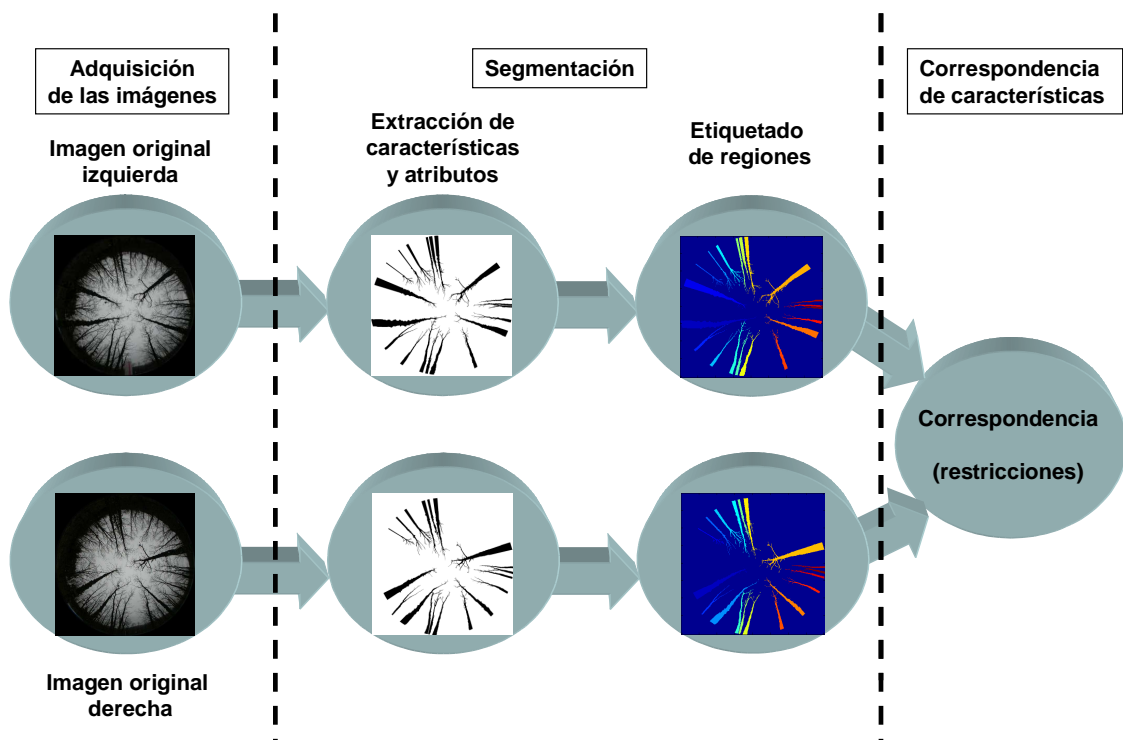


Figura 3.2 Esquema del proceso de segmentación y correspondencia basado en regiones.

### 3.3.2 Extracción de características

Como se ha mencionado reiteradamente, ahora las características son regiones, cuyo proceso de extracción se fundamenta en los siguientes pasos, que se aplican de forma secuencial:

**Paso 1 *Imagen válida*:** el CCD de la cámara utilizada posee  $2272 \times 1704$  píxeles de ancho y largo respectivamente, de los que solamente se utilizan  $1616 \times 1616$ , que son las dimensiones que contienen al círculo central, tal y como se explicó en el capítulo uno. Por tanto, considerando el origen de coordenadas en la esquina superior izquierda, el centro de la imagen se localiza en las coordenadas (808, 808). El radio  $R$  de la imagen válida desde el centro es de 808 píxeles. De este modo, durante el proceso sólo se considera la región de la imagen dentro del área limitada por el radio dado. Además, trabajamos con la imagen de intensidad  $I$  en el espacio de color  $HSI$  obtenido tras la transformación desde el modelo  $RGB$ . Esto es así porque, como se explicó en el capítulo uno, no hemos alcanzado resultados satisfactorios con los espacios de color estudiados, mientras que la imagen  $I$  contiene la información espectral de los tres canales espectrales  $R$ ,  $G$  y  $B$ . Por otro lado y también tal y como se explicó en el capítulo uno, existía la posibilidad de aplicar el proceso de segmentación sobre las imágenes realzadas pertenecientes al tipo mostrado en la figura 1.6. Sin embargo, el proceso de crecimiento de regiones o “*region growing*”, que se aplica posteriormente se ha comprobado que es más eficaz sobre las imágenes originales, mostradas en la figura 1.5, que sobre las realzadas, figura 1.6, debido principalmente al mayor grado de similitud en los valores de intensidad de las originales. Esto justifica el uso de las imágenes originales en lugar de las realzadas. Más adelante, en el capítulo cinco, se proporcionan los detalles acerca del protocolo para la medición de una parcela en el bosque, donde el sensor se sitúa en el centro de un círculo con un radio que varía desde  $5m$  a  $25m$ . Sólo los troncos dentro de estos círculos con radio de  $25m$  son realmente los de interés. Además, un hecho relevante a este respecto es que los troncos por debajo de la distancia de  $25m$  se proyectan en la imagen con un área suficiente para su posterior tratamiento, mientras que los situados a distancias superiores a los  $25m$  se proyectan con áreas mucho menores, haciendo que en la mayoría de los casos su procesamiento pueda llegar a ser complicado debido a su reducido tamaño respecto del área.

**Paso 2** *Circunferencias concéntricas*: dibujamos circunferencias concéntricas sobre la imagen original, comenzando con un radio  $r$  de tamaño igual a 250 píxeles y con incrementos de 50 píxeles hasta  $r = R$ . Para cada circunferencia obtenemos su perfil de intensidad. Existen principalmente dos tipos de circunferencias, a saber: aquéllas en las que los troncos atravesados se manifiestan perfectamente contrastados con respecto al fondo y aquéllas que atraviesan zonas donde el fondo y los troncos se confunden. La figura 3.4(a) muestra ambos tipos de circunferencias, las primeras se dibujan en amarillo y las segundas en rojo. Obsérvese a este respecto, cómo las circunferencias amarillas cruzan áreas con los troncos sobre el cielo y las rojas cruzan zonas donde los troncos y el suelo aparecen con niveles de intensidad similares.

**Paso 3** *Perfiles de intensidad y detección de regiones oscuras*: siguiendo los recorridos de las circunferencias, dibujamos los perfiles de intensidad asociados a cada una de ellas. La figura 3.3 muestra dos perfiles de intensidad cubriendo un rango de  $45^\circ$ , concretamente de  $135^\circ$  a  $180^\circ$ . En la gráfica del perfil aparecen niveles de intensidad bajos y altos. Los niveles bajos se asocian a los troncos o al suelo y los altos al cielo. Basado en lo anterior, si aparecen grandes zonas oscuras en el perfil, esto significa que la circunferencia cruza una región donde los troncos y el suelo no pueden distinguirse, en cuyo caso la circunferencia generada por el píxel se etiqueta en rojo. Esto ocurre en la figura 3.3(a) que representa valores de intensidad bajos, variando de 0 a 0.18 sobre el rango  $[0,1]$ , es decir, una zona oscura extensa. Si no se identifican grandes zonas oscuras, las circunferencias se etiquetan en amarillo. Por el contrario, si un área oscura relativamente pequeña aparece limitada por dos áreas claras, se asume que el perfil generado por la circunferencia atraviesa un tronco, figura 3.3(b).

El concepto de área extensa o reducida se establece en función de un valor de umbral que se fija en el paso 5, de suerte que las áreas oscuras con valores por encima de dicho umbral se consideran formadas por la fusión de dos regiones, una de ellas el suelo y la otra probablemente los troncos.

**Paso 4** *Siembra de semillas en los troncos*: considerando las circunferencias amarillas, podemos detectar las posiciones de los troncos atravesadas por ellas, que son regiones oscuras homogéneas en el perfil limitado por zonas claras, figura 3.3(b). Esto permite escoger un píxel para cada región homogénea oscura; este píxel se considera a partir de este momento como una semilla que marcará el inicio del proceso de

crecimiento posterior. Además, porque sabemos que la transición de claro a claro atraviesa una región homogénea oscura, obtenemos su valor de intensidad medio y desviación estándar. En resumen, dada una región homogénea oscura definida por una circunferencia amarilla, sobre dicha región se selecciona un píxel que hace las funciones de semilla y sobre los píxeles atravesados por dicha circunferencia se obtiene su valor medio y desviación estándar, como dos atributos asociados a dicha semilla.

**Paso 5 Filtrado de la región:** solamente estamos interesados en unas regiones oscuras específicas, aquéllas que representan los troncos de los árboles. El proceso para elegir las zonas que potencialmente pueden representar troncos, figura 3.4(b), es el siguiente:

- a) Consideramos solamente aquellas regiones oscuras en el perfil donde la intersección con circunferencias amarillas producen una línea con más de  $U_1$  píxeles. Esto garantiza que los troncos analizados sean suficientemente anchos. Su justificación se debe al hecho de que se asume que este tipo de troncos pertenecen al área de interés objeto de análisis, es decir, el círculo con radio menor de  $25m$ .
- b) Asimismo, sobre la base de las circunferencias amarillas, sólo tenemos en cuenta las regiones con niveles de intensidad menores que  $U_2$ , porque estamos tratando con regiones homogéneas oscuras (troncos).
- c) Considerando la circunferencia más externa de color amarillo  $c_i$ , seleccionamos únicamente las regiones oscuras cuya intersección con esta circunferencia genere una línea en el perfil con un número de píxeles menor que  $U_3$ . El valor máximo en píxeles de todas las líneas de intersección es  $t_{\max}^i < U_3$ . Entonces para la siguiente circunferencia amarilla hacia el centro de la imagen,  $c_{i+1}$ ,  $U_3$  toma el valor de  $t_{\max}^i$ , que es el valor utilizado cuando se procesa la siguiente circunferencia, y así sucesivamente hasta llegar a la circunferencia más interna. Esto se justifica por el hecho de que el grosor de los troncos siempre disminuye hacia el centro de la imagen.

En este trabajo,  $U_1$ ,  $U_2$  y  $U_3$  se fijan a 10, 0.3 y 120 respectivamente, tras diversas pruebas experimentales basadas en ensayo y error. Conviene señalar a este

respecto que el valor umbral  $U_2$ , que representa niveles de intensidad, se refiere al rango  $[0,1]$  en lugar de la habitual representación de los valores de intensidad con 8 bits y por tanto  $[0, 255]$ , lo que equivale a un valor de intensidad de 76 con respecto a este último rango.

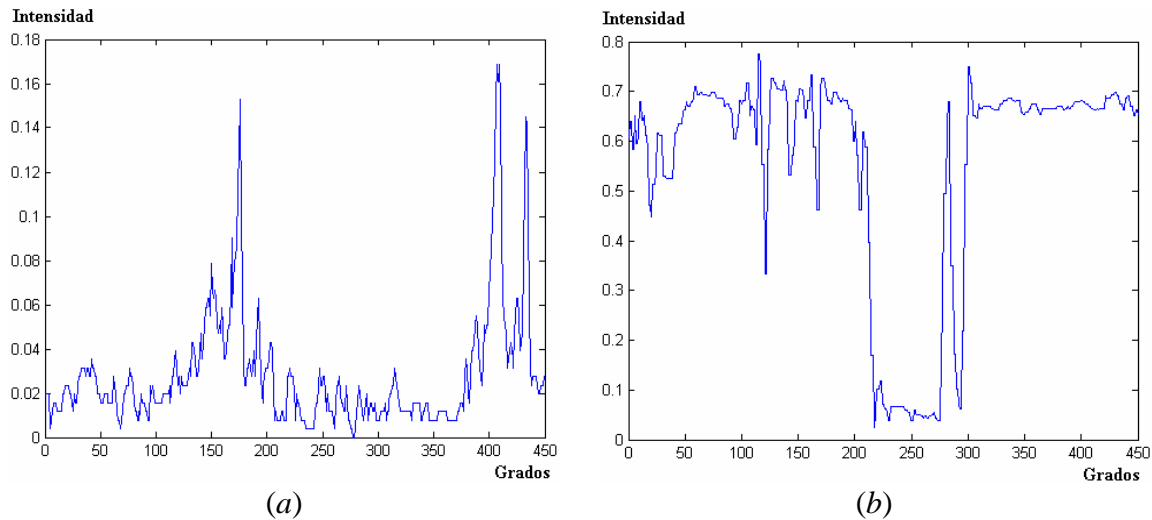


Figura 3.3 Perfiles de intensidad. (a) La circunferencia atraviesa una región donde los troncos y el suelo no pueden distinguirse. (b) Niveles de intensidad bajos y altos. El primero se asocia a troncos y el segundo al cielo. Las intensidades varían de 0 a 1.

**Paso 6 Crecimiento de regiones:** este proceso se basa en el procedimiento descrito en Gonzalez y Woods (2008). Comenzamos en la circunferencia amarilla más externa, seleccionando los píxeles semilla obtenidos para esta circunferencia. A partir de estas semillas añadimos a cada una de ellas aquellos píxeles vecinos que poseen un valor de intensidad similar al correspondiente atributo de la semilla establecida en el paso 4. La similitud se mide como la diferencia entre el valor de intensidad del píxel objeto de consideración y el valor medio en la zona a la que la semilla pertenece, sin diferir más de la desviación estándar también calculada en el paso 4 para esa zona. El crecimiento de regiones finaliza cuando no se encuentran más píxeles vecinos similares para esa semilla entre esta circunferencia y el centro de la imagen. Esto permite obtener un conjunto de regiones como el mostrado en la figura 3.4(c).

**Paso 7 Etiquetado:** como una fase previa al etiquetado, se aplica una operación morfológica de apertura. El objetivo consiste en romper enlaces poco significativos entre regiones, para evitar que algunas ramas de los árboles superpuestas con otras ramas o troncos lleve a etiquetar dos árboles o troncos como una única región. El elemento estructural durante la apertura es la clásica estructura de dimensión  $3 \times 3$  con

1's en todos sus elementos. Se elige este elemento por sus propiedades de simetría, siendo capaz de operar con la misma fuerza en todas las direcciones espaciales de la imagen, esto es, en horizontal, vertical y diagonal.

Las regiones extraídas durante el crecimiento de regiones anterior se etiquetan siguiendo el algoritmo de etiquetado de componentes conexas descrito en Haralick y Shapiro (1992). La figura 3.4(d) muestra el resultado del etiquetado obtenido cuando se aplica dicho algoritmo a la imagen mostrada en la figura 3.4(c), donde cada color representa el valor numérico de la etiqueta.

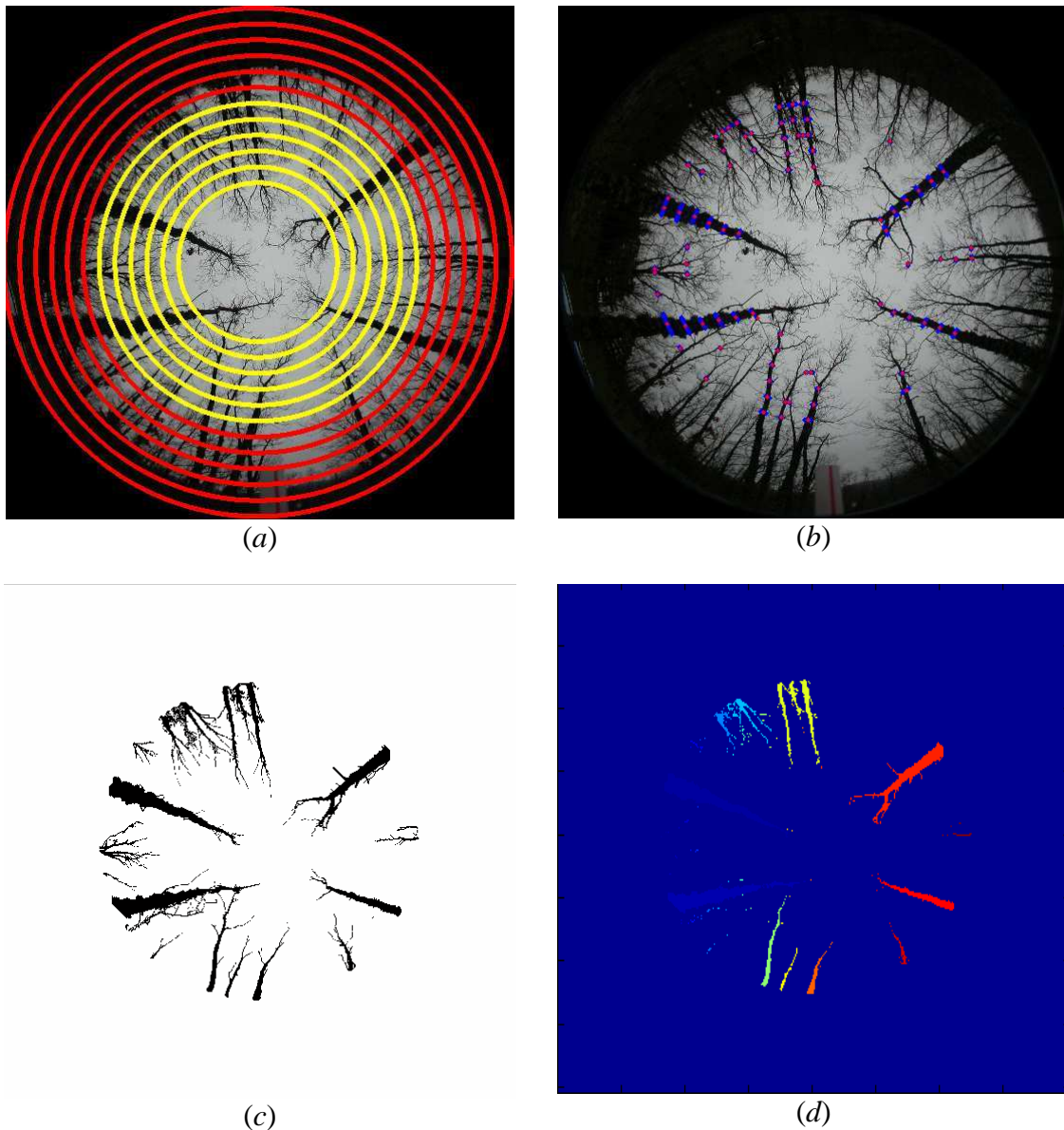


Figura 3.4 (a) Circunferencias dibujadas en amarillo y rojo. (b) Regiones homogéneas dibujadas en azul y las semillas en rojo. (c) Imagen resultante obtenida con el proceso de *crecimiento de regiones*. (d) Imagen resultante obtenida mediante el proceso de *etiquetado*.

**Paso 8 Asociación de regiones y semillas:** para cada una de las semillas en la circunferencia amarilla más externa, hacemos corresponder su región identificada previamente. Es posible que más de una semilla pertenezca a la misma región. Si esto ocurre, creamos nuevas regiones, de modo que finalmente se obtenga el mismo número de regiones que de semillas. Tras este paso, cada región tiene asignada una única semilla.

**Paso 9 Asociación de semillas:** seleccionamos las otras semillas disponibles en el resto de circunferencias amarillas. Si una de estas semillas cumple que es la más próxima en términos de distancia de píxel, y su ángulo en grados el más similar al ángulo de la semilla seleccionada previamente, entonces se asume que pertenece a la misma región que la semilla anterior, que hace de referencia. El ángulo en grados es el valor  $\theta$  en coordenadas polares  $(\rho, \theta)$  con respecto a la localización de la semilla en coordenadas cartesianas  $(x, y)$ . Este proceso permite establecer correspondencias entre las semillas de las diferentes circunferencias amarillas dependiendo de la región a la cual pertenece, figura 3.5(a), es decir, se trata de identificar semillas que pertenecen probablemente a la misma región (tronco). Calculamos la orientación media,  $\bar{s}$ , para todas las semillas pertenecientes a la misma región identificada de acuerdo con el proceso descrito en este punto. La orientación media se refiere exactamente al valor medio de los ángulos  $\theta$  para todas las semillas.

**Paso 10 Estimación de la localización de las semillas en las circunferencias rojas:** este proceso consta a su vez de tres subpasos, *predicción*, *corrección* y *medición*:

- a. *Predicción:* los píxeles pertenecientes a un tronco atravesado por una circunferencia roja deben tener idéntica orientación, en unidades de grados, que la semilla en la circunferencia amarilla más externa que atraviesa el mismo tronco. De este modo obtenemos las semillas en las circunferencias rojas que cumplen la anterior restricción, comenzando desde la circunferencia más interna.
- b. *Corrección:* puesto que existen troncos que no están alineados hacia el centro, la *predicción* puede introducir errores. Se aplica un factor de corrección (“*offset*”) a esta posición, que resulta ser exactamente  $\bar{s}$ , calculado en el paso 9, figura 3.5(b).



- c. *Medición*: tras aplicar el factor de corrección, verificamos si la semilla estimada en cada circunferencia roja pertenece a un tronco. Esto sólo es posible si la circunferencia roja atraviesa una región con valores de intensidad bajos y además está limitada por zonas con valores de intensidad altos y la posición de la semilla estimada está dentro de la región con valores bajos. Con esto asumimos que la semilla pertenece al mismo tronco que la semilla en la circunferencia amarilla. Debido al contraste en el perfil de intensidad para la región roja en el tronco específico, podemos medir la posición exacta de la semilla en la parte central de la región con baja intensidad. La posición estimada de la semilla se sustituye por la medida y se utiliza nuevamente para estimar la posición de la próxima semilla en la siguiente circunferencia roja. Si el perfil no presenta valores de intensidad bajos y altos, no pueden tomarse mediciones y la localización de la siguiente semilla es la estimada previamente por la *predicción* y *corrección*.

**Paso 11** *Nuevo crecimiento de regiones*: comenzando con la circunferencia amarilla más externa, aplicamos de nuevo un proceso de crecimiento de regiones como el descrito en el paso 6, pero ahora controlado por varias iteraciones (tantas como circunferencias rojas existan). Para cada iteración, el crecimiento de regiones tiene su límite superior dado por el radio de la circunferencia roja más próxima.

Una vez se alcanza la circunferencia roja más externa, es decir, el máximo número de iteraciones, el proceso de crecimiento de regiones finaliza. En este momento se aplica una operación morfológica de apertura que intenta romper enlaces entre regiones (troncos) que pudieran estar todavía unidos. El elemento estructural empleado en la apertura es el mismo que el utilizado en el paso 7. La figura 3.5(c) muestra el resultado de aplicar este paso sobre la imagen representada.

**Paso 12** *Re-etiquetado*: este proceso es similar al descrito en el paso 7. Se vuelven a etiquetar cada una de las regiones que han aparecido después del proceso de crecimiento de regiones del paso 11, figura 3.5(d).

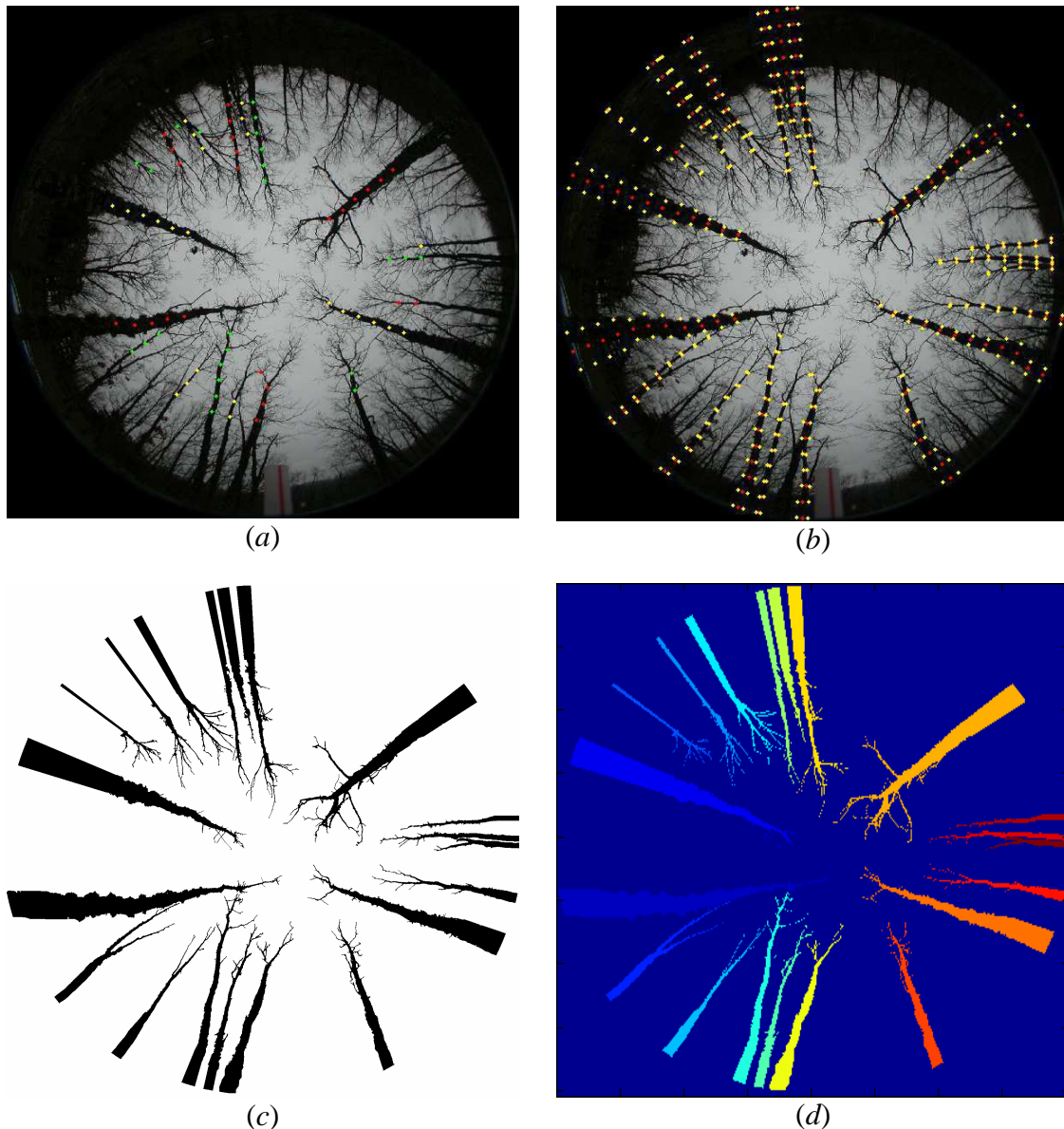


Figura 3.5 (a) Semillas clasificadas dependiendo de la región a la cual pertenecen. Por claridad podemos ver las semillas pintadas con el mismo color. (b) Imagen resultante obtenida con los pasos: *predicción* y *corrección*. Las semillas se pintan en rojo. Cada región está delimitada con puntos amarillos teniendo en cuenta las circunferencias amarillas asociadas. (c) Imagen resultante obtenida con el segundo proceso de *crecimiento de regiones*. (d) Imagen resultante obtenida después del proceso de *re-etiquetado*.

### 3.3.3 Extracción de atributos

**Paso 13 Extracción de atributos:** una vez que todas las regiones han sido etiquetadas de nuevo, para cada región extraemos los siguientes atributos: *a) área* (número de píxeles), *b) centroide* (posiciones  $xy$  del píxel medio en la región), *c) ángulo* en grados de cada centroide según el sistema de coordenadas polares con origen en el centro de la imagen y *d) los siete momentos invariantes de Hu* (Gonzalez y Woods, 2008; Pajares y Cruz, 2007a,b).

Para el cómputo de los momentos invariantes se parte de los valores  $f(x,y)$ , que representan las intensidades de los píxeles  $(x,y)$  en la región bajo consideración. El momento de orden  $(p + q)$  para la región se define como:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (3.16)$$

donde el sumatorio se toma sobre todas las coordenadas espaciales  $(x,y)$  de píxeles de la región. El *momento central* de orden  $(p + q)$  viene dado por:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (3.17)$$

donde

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (3.18)$$

Los momentos *centrales normalizados* de orden  $(p + q)$  se definen como:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \text{donde} \quad \gamma = \frac{p+q}{2} + 1 \quad \text{para } (p + q) = 2, 3, \dots \quad (3.19)$$

Finalmente, los siete *momentos invariantes de Hu* se obtienen a partir de los momentos centrales normalizados de órdenes 2 y 3 como sigue:

$$\phi_1 = \eta_{20} + \eta_{02} \quad (3.20)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (3.21)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (3.22)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (3.23)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left[ (\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left[ 3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \quad (3.24)$$

$$\phi_6 = (\eta_{20} - \eta_{02}) \left[ (\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (3.25)$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (3.26)$$

La principal ventaja de estos momentos es que son invariantes a rotaciones, traslaciones y cambios de escala, circunstancia ésta que se aprovecha en nuestro diseño, dado que las regiones en las imágenes aparecen rotadas y con ligeras variaciones en lo que a la escala se refiere, tal y como se verá en capítulos posteriores.

### 3.3.4 Resumen del proceso de segmentación basado en regiones

Tal y como se ha expuesto previamente, el proceso de segmentación basado en regiones consta de lo que podríamos denominar una primera fase relativa a la extracción de regiones, que se utilizan como características, según los pasos 1 a 12, seguida de una segunda fase donde se obtienen las propiedades de dichas regiones.

En la tabla 3.2 se muestran los tres umbrales involucrados en el proceso de segmentación de regiones.

Tabla 3.2 Umbrales implicados en el proceso de segmentación basado en regiones

Fase	Parámetro/Umbrales	Descripción
Extracción de regiones	$U_1 = 10$ $U_2 = 0.3$ $U_3 = 120$	Parámetros involucrados en el paso cinco, relativo al filtrado de las regiones



## Correspondencia estereoscópica

### 4.1 Introducción

Como se ha mencionado en el capítulo dos y reiterado específicamente en el tres, los pasos esenciales del proceso de visión estereoscópica abordados en este trabajo son por un lado la segmentación, mediante la cual se extraen las características de las imágenes conjuntamente con sus atributos, y por otro lado la correspondencia de esas características. En el capítulo tres se abordó la parte relativa a la segmentación desde una doble vertiente. Corresponde centrarnos en este capítulo concretamente en la correspondencia de dichas características.

Siguiendo los dos enfoques mantenidos hasta ahora, tratamos la correspondencia de características desde ambos puntos de vista, a saber: correspondencia basada en píxeles y correspondencia basada en regiones. La primera se trata en la sección 4.2 y la segunda se aborda en la 4.3.

Al tratarse de una correspondencia estereoscópica el problema se enfoca desde la aplicación de una serie de restricciones comúnmente utilizadas en cualquier proceso de esta naturaleza, que han sido introducidas en la sección 2.3.1 y particularizadas para el tipo de imágenes a tratar en la presente investigación, que considera la problemática derivada de las mismas.

De forma general y según las restricciones aplicadas, el objetivo consiste en establecer las correspondencias entre características de forma inequívoca, considerando naturalmente las particularidades derivadas de la naturaleza del sistema estereoscópico, tales como diferentes proyecciones de una misma estructura 3D en las dos imágenes, donde las oclusiones y otros fenómenos hacen acto de presencia (Pajares y Cruz, 2007a; Ayache 1991). Por la importancia que las restricciones estereoscópicas tienen en cualquier sistema estereoscópico y por tanto en el que nos compete, a continuación ampliamos y precisamos los aspectos más relevantes existentes entre las restricciones y nuestro sistema estereoscópico. Las restricciones de epipolaridad, semejanza y unicidad se aplican en los dos enfoques abordados en este capítulo, mientras que la de orden posicional se utiliza en exclusiva en el enfoque basado en regiones, y la de continuidad de la disparidad en el basado en píxeles.

- a) *Epipolar*: las imágenes de una misma entidad física en el espacio 3D deben proyectarse sobre la misma línea epipolar. Las líneas epipolares en los sistemas convencionales son generalmente horizontales, que intersectan ambas imágenes a la misma altura. En el caso del sistema de medición que proporciona nuestras imágenes estereoscópicas forestales, dichas líneas vienen delimitadas por arcos de circunferencia que recorren las imágenes en sendos hemisferios, izquierdo y derecho, con diferentes radios, según la ubicación de los píxeles con respecto a los centros de las imágenes. Los detalles sobre la obtención de las líneas epipolares se proporcionan en la sección 4.2.2. El procedimiento de obtención sólo depende del sistema de medición, su geometría y la forma de proyección, basada en la óptica con lente de ojo de pez. Por tanto se trata de las mismas líneas, independientemente del tipo de imágenes capturadas por el sistema.
- b) *Semejanza*: las dos proyecciones, identificadas en las imágenes como características, de la misma entidad 3D deben tener propiedades o atributos similares. Como se ha mencionado previamente, esas características son píxeles o regiones, dependiendo del tipo de imágenes. Por tanto, para el caso de la correspondencia basada en píxeles serán los atributos descritos en la sección 3.2.4 los utilizados en la aplicación de esta restricción, esto es, coeficiente de correlación, color, textura, módulo y dirección del gradiente y Laplaciana; mientras que en el caso de la correspondencia basada en regiones los atributos utilizados en esta restricción son los descritos en la sección

3.3.3, es decir, área, centroide, ángulo del centroide y los siete momentos invariantes de Hu.

- c) *Unicidad*: para cada característica en una imagen debe haber una única característica en la otra imagen. Se trata de una restricción que requiere una especial consideración en el caso del enfoque basado en píxeles propuesto en este trabajo. Como se verá posteriormente, debido a la geometría del sistema estereoscópico, puede haber píxeles en una determinada imagen con más de una correspondencia en la segunda imagen del par estereoscópico. Incluso, debido a la distribución de los troncos en las imágenes pueden aparecer oclusiones de unos troncos frente a otros, lo que implica que no se encuentren las correspondencias verdaderas. Ambos factores, se tienen en cuenta de forma especial en la estrategia propuesta para el enfoque de correspondencia basado en píxeles.
- d) *Orden posicional*: en un sistema estereoscópico convencional de ejes ópticos paralelos, dadas dos características en una determinada imagen, por ejemplo la izquierda, situada una a la derecha de la otra, esta restricción supone que este mismo orden se mantiene en la imagen derecha para sus respectivas características homólogas. Así, dado el par de imágenes estereoscópicas de la figura 4.1(a), si suponemos que las características  $i$  y  $h$  de la imagen izquierda se corresponden con la  $j$  y  $k$  respectivamente en la imagen derecha, en esta situación se verificará que  $x_i < x_h$  y  $x_j < x_k$ . En el caso de nuestras imágenes, cuyo fundamento deriva de la óptica basada en lentes de ojo de pez, la restricción epipolar tiene la misma filosofía que la de los sistemas convencionales, si bien ahora las posiciones relativas se mantienen respecto de los ángulos. En efecto, dado el par de imágenes estereoscópicas de la figura 4.1(b), como en el caso anterior, si suponemos que las características  $i$  y  $h$  de la imagen izquierda se corresponden con la  $j$  y  $k$  respectivamente en la imagen derecha, en esta situación se verificará que  $\theta_i < \theta_h$  y  $\theta_j < \theta_k$ , donde  $\theta_i, \theta_h, \theta_j, \theta_k$  representan el valor de los ángulos correspondientes de los píxeles en coordenadas polares.

En cualquier caso, es necesario tener en cuenta ciertas precauciones a la hora de aplicar esta restricción, en particular en lo que respecta a situaciones donde la misma puede ser fácilmente violada, tal y como ocurre en algunos casos descritos en la



sección 4.3.4. En cualquier caso, se trata de una restricción con una fuerte dependencia de la restricción epipolar.

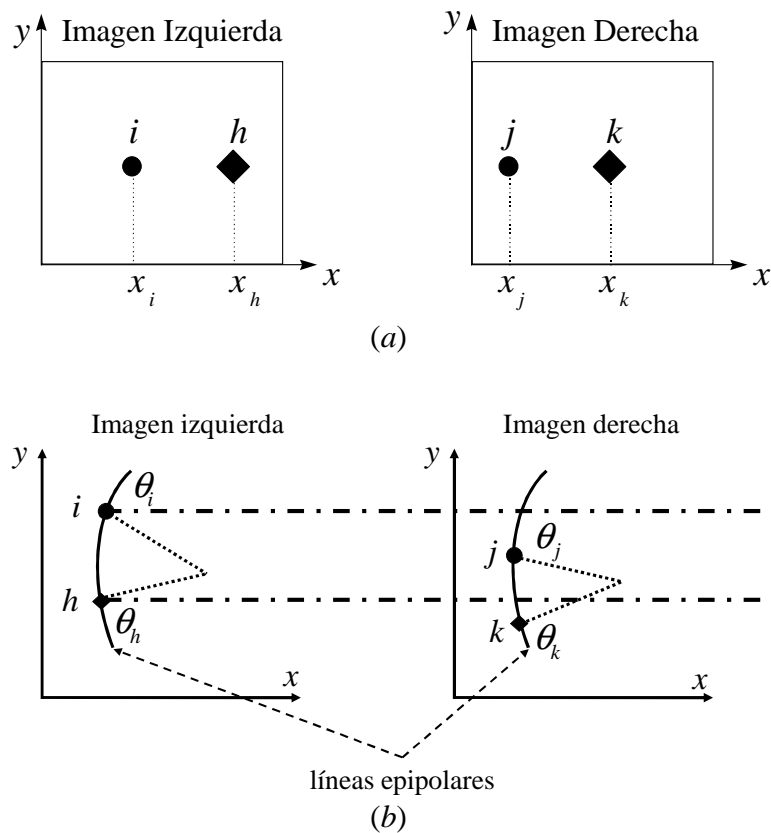


Figura 4.1 Restricción de orden posicional: (a) sistema estereoscópico convencional basado en proyección de perspectiva con los ejes ópticos paralelos; (b) sistema basado en lentes de ojo de pez con ejes ópticos paralelos.

- e) *Continuidad de la disparidad*: asume que las variaciones de disparidad en la imagen son generalmente suaves, es decir, que si consideramos un mapa de disparidad éste se presenta continuo salvo en las zonas que presentan discontinuidades estructurales, como ya se ha mencionado en la sección 2.3.1. Esta restricción también aparece bajo distintas formas y a veces con alguna pequeña variación en algunas referencias bibliográficas, tal es el caso de *Disparidad Diferencial Mínima* en Medioni y Nevatia (1985) y Pajares y col. (1998) o *Continuidad “figural”* en Pollard y col. (1981). Esta restricción requiere una consideración especial en el caso de nuestras imágenes forestales, ya que la misma sólo es aplicable en zonas donde dicha discontinuidad se mantiene, lo cual ocurre exactamente en el interior de los troncos, incumpléndose en la mayoría de las zonas de la imagen y en particular en los bordes mismos de los troncos, dado que por detrás de ellos las disparidades pueden ser de lo más diversas. Como quiera que esta restricción implica que dada una característica con un cierto

nivel de disparidad, en sus proximidades se asume la existencia de otras características con niveles de disparidad similares, es evidente que esto sólo es posible exactamente dentro de los troncos de los árboles y por tanto aplicable a los píxeles en el interior de los mismos, perdiendo la razón de su aplicabilidad para el resto de estructuras y en particular en los bordes mismos de los troncos. Por este motivo esta restricción sólo es factible en el caso del enfoque basado en píxeles y no en el de regiones.

## 4.2 Correspondencia basada en píxeles

### 4.2.1 Restricciones estereoscópicas: diseño del proceso

De acuerdo con el esquema de la figura 3.1, una vez que el proceso de segmentación ha finalizado, se han identificado píxeles tanto en la imagen izquierda como en la derecha pertenecientes a los tres tipos de texturas que van a ser descartados durante el proceso de correspondencia estereoscópica por carecer de interés en este tipo de imágenes, tal y como se explicó en los capítulos precedentes. El proceso de correspondencia se lleva a cabo en paralelo entre las imágenes izquierda y derecha. Por tanto, en primer lugar se toma como referencia la imagen izquierda donde se establecen las pertinentes correspondencias con la imagen derecha, si existen, es decir cuando las oclusiones lo permitan. En segundo lugar se realiza el mismo proceso tomando como referencia la imagen derecha. En ambos casos el proceso de correspondencia estereoscópica sólo se aplica sobre los píxeles que no fueron excluidos, entre los que deben encontrarse los que pertenecen a los troncos.

Las restricciones estereoscópicas anteriormente descritas, que se aplican sobre este tipo de imágenes por las razones expuestas, son las siguientes: epipolar, semejanza, unicidad y continuidad de la disparidad.

La restricción epipolar permite restringir el espacio de búsqueda para la correspondencia entre píxeles, ya que la geometría del sistema y la proyección basada en la óptica de “ojo de pez” así lo determina. En la sección 4.2.2 se detalla el procedimiento para su aplicación.

La restricción de semejanza permite establecer medidas de similitud entre pares de píxeles procedentes de la imagen izquierda con píxeles de la imagen derecha y

viceversa, de derecha a izquierda. De esta forma, para cada píxel de una de las imágenes se dispone de un conjunto de píxeles en la otra imagen, que constituyen una lista de posibles candidatos. El objetivo final es elegir de entre los candidatos de la lista uno de ellos como el preferido. Llegados a este punto, todavía restan por aplicar las restricciones de continuidad de la disparidad y unicidad. Dependiendo del orden en el que se apliquen a partir de este momento, caben dos alternativas de diseño posibles para completar el proceso de correspondencia estereoscópica, a saber:

- 1) aplicar unicidad seguida de continuidad, figura 4.2, rama izquierda
- 2) aplicar continuidad y posteriormente unicidad, figura 4.2, rama derecha.

En el primer caso, una vez que se dispone de un conjunto de medidas de similitud entre píxeles de la imagen izquierda con píxeles de la derecha, se procede a tomar una decisión sobre cuál es la correspondencia considerada como verdadera. Esto se puede llevar a cabo por cualquier mecanismo de decisión, teniendo en cuenta las preferencias en ambos sentidos (izquierda a derecha y derecha a izquierda), de forma que mediante la aplicación de la restricción de unicidad, se elige un único candidato de entre todos los posibles. De esta forma se obtiene un mapa de disparidad inicial. Lo más probable es que dicho mapa contenga errores, incluyendo falsos positivos (Rohith y col., 2008). La idea principal a partir de aquí consiste en aplicar el concepto de continuidad de la disparidad, de forma que siguiendo los principios de la Gestalt (Wang, 2005), se asume que las disparidades de píxeles vecinos son similares excepto en determinadas zonas de discontinuidad, que se presentan precisamente de forma más acusada en los bordes de los árboles. En esta línea se han utilizado distintas técnicas tales como filtros bilaterales (Ansar y col., 2004), ajustes de planos sobre los mapas de disparidad (Klaus y col., 2006) o filtros estadísticos como media, mediana o moda (Lankton, 2010). En cualquier caso, todas estas técnicas aplican la restricción de continuidad de la disparidad, tras lo cual se obtiene un nuevo mapa de disparidad refinado, donde determinados valores de disparidad espurios o erróneos se consiguen eliminar.

En el segundo caso, tras la aplicación de la restricción de semejanza, no se toma todavía una decisión encaminada a seleccionar emparejamientos correctos, antes bien, se opta por aplicar la restricción de continuidad sobre la base de que dado un píxel en la

imagen izquierda y su homólogo en la derecha, los píxeles vecinos de ambos, en sendas imágenes, deben ser también homólogos entre sí, asumiendo que la disparidad es continua en dichas vecindades. Estos procesos suelen ser de optimización basados en la minimización de una función de energía (Banno e Ikeuchi, 2009; Ruichek y Postaire, 1996), de suerte que una vez finalizado el proceso de optimización, los píxeles han modificado los valores iniciales de similitud, adquiriendo nuevos valores, en base a los cuales y por aplicación de la restricción de unicidad se obtiene el mapa de disparidad. Para ello, como en el caso del mapa inicial, se aplica algún proceso de decisión.

En cualquiera de los enfoques, la restricción de unicidad merece una especial atención, debido a las particularidades derivadas de la geometría del sistema como se verá posteriormente.

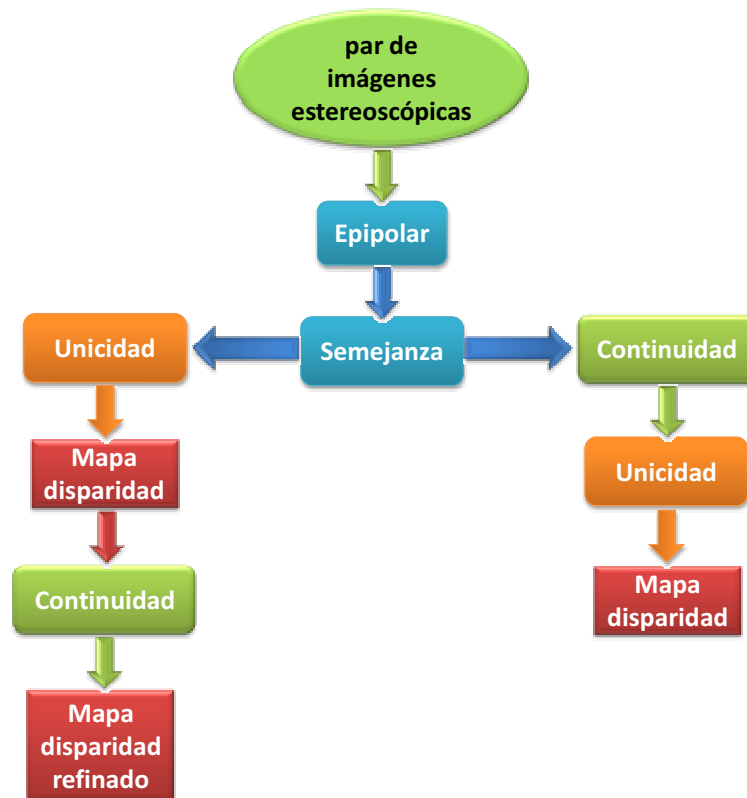


Figura 4.2 Opciones en la aplicación de restricciones para la obtención del mapa de disparidad

La cuestión que se plantea ahora consiste en decidir qué estrategia utilizar. La decisión finalmente adoptada es la primera, es decir la que sigue la rama izquierda de la figura 4.2, por las razones expuestas a continuación. Tras un análisis exhaustivo de ambos procedimientos se llega a la conclusión de que los mejores resultados se obtienen en el primer caso, lo cual proviene de la naturaleza misma de las imágenes. En efecto,

se trata de imágenes en las que llegados a este punto, y aún habiendo excluido del proceso de correspondencia las regiones mencionadas durante el proceso de segmentación (hojas, cielo, hierba del suelo), todavía persiste un importante número de píxeles con un alto índice de violación de la restricción de continuidad de la disparidad. En efecto, son aquellos píxeles situados principalmente fuera de los troncos de los árboles y sobre los bordes mismos de dichos troncos. Como se ha mencionado previamente, la aplicación de la restricción de continuidad de la disparidad sobre estos píxeles, indudablemente conduce a suposiciones erróneas por el mero hecho de considerar que un píxel y sus vecinos poseen disparidades idénticas y por consiguiente a resultados incorrectos. Este hecho se acentúa en los bordes de los troncos, sobre los que se tomarían decisiones incorrectas con bastante certeza, algo a todas luces indeseable dado que los troncos constituyen el objetivo principal de nuestro análisis forestal.

En resumen, el incumplimiento de la restricción de continuidad, en las situaciones mencionadas, hace que la aplicación de dicha restricción, conjuntamente con la de similitud como paso previo a la decisión, introduzca errores en la actualización de los valores de similitud, conduciendo finalmente a decisiones incorrectas y a un mapa de disparidad con errores al menos en las posiciones de los píxeles que no cumplen la restricción de continuidad. Por consiguiente, en este trabajo de investigación se ha apostado por aplicar un proceso de decisión lo más fiable posible para determinar las correspondencias de manera local, basándonos exclusivamente en la similitud píxel a píxel, sin tener en cuenta en este momento las vecindades de los mismos, de suerte que se obtiene un mapa de disparidad a partir de dichas decisiones por aplicación de la restricción de unicidad. Aplicando sobre este mapa la restricción de continuidad de la disparidad, lo que se consigue es no modificar los valores de similitud, que son determinantes para la decisión, sino realizar las modificaciones sobre los valores de disparidad mismos, sin necesidad de tomar decisiones posteriores. En resumen, son estas razones las que justifican la opción de diseño elegida para el proceso de correspondencia. Además, algunos experimentos preliminares ya apuntaban hacia mejores resultados por parte de la opción elegida, lo que de algún modo confirma lo adecuado de la elección.

Como se ha mencionado previamente, en el proceso de correspondencia, un elemento clave es el mecanismo de decisión en base al cual se obtiene el mapa inicial de

disparidad. Por este motivo se han estudiado los siguientes enfoques, que son una adaptación de técnicas de clasificación bajo el paradigma basado en la toma de decisiones, a saber:

- a) Integral Fuzzy de Choquet o “*Choquet Fuzzy Integral*” (CFI), descrito en Herrera y col. (2009a).
- b) Integral Fuzzy de Sugeno o “*Sugeno Fuzzy Integral*” (SFI), descrito en Herrera y col. (2009b).
- c) Teoría de Dempster-Shafer (DES), descrito en Herrera y col. (2009b).
- d) Teoría de la Decisión Multicriterio Fuzzy o “*Multicriteria Decision Making Fuzzy*” (MCDMF), descrito en Herrera y col. (2009c).

Todos ellos, como se ha indicado previamente, son en realidad métodos basados en la teoría de toma de decisiones, de suerte que dado un píxel de referencia con sus atributos asociados, por ejemplo en la imagen izquierda, se buscan posibles candidatos en la imagen derecha, cada uno con sus propias características, y viceversa, para un píxel de referencia en la imagen derecha y sus candidatos en la izquierda. Entre el píxel de referencia y los candidatos se establecen una serie de medidas de similitud, en base a las cuales se elige el mejor o mejores candidatos mediante un criterio basado en la toma de decisiones involucrada. Las medidas de similitud provienen de la aplicación de la restricción del mismo nombre. Conviene destacar en este punto el hecho, ya referido, de que a un píxel puede corresponderle más de un candidato en la otra imagen por las especiales características del sistema de medición, cuestión ésta que se aborda en la sección 4.2.4.

Tras la aplicación de las tres restricciones mencionadas, epipolar, semejanza y unicidad, se obtiene un mapa de disparidad, que denominamos *inicial*, tomando como referencia la imagen izquierda. En este mapa los píxeles excluidos durante el proceso de segmentación, por pertenecer a una de las tres texturas carentes de interés, se etiquetan con disparidad cero, como se explica posteriormente, el resto con la disparidad que les asigna el proceso de decisión en cuestión según el grado de similitud que poseen en relación a sus candidatos.

Una vez obtenido el mapa de disparidad inicial, el objetivo se centra ahora en elegir un mecanismo apropiado para filtrar dicho mapa mediante la aplicación de la restricción de la continuidad de la disparidad. Tanto los procesos que aplican un filtrado de naturaleza estadística o ajustan planos a los valores de disparidad obtenidos, son procesos globales que consideran el conjunto de píxeles en la región de donde extraen los valores. En la búsqueda de la mejor solución, lo más conveniente consiste en considerar las contribuciones de los píxeles vecinos de forma individualizada. Es decir, cada píxel vecino contribuirá de modo diferente a la modificación de los valores de disparidad en función de su relación con el píxel que se está tratando de modificar su disparidad. En esta línea se sitúan las soluciones basadas en el paradigma de los denominados Mapas Cognitivos Fuzzy o “*Fuzzy Cognitive Maps*” (*FCMs*) y la Red Neuronal de Hopfield o “*Hopfield Neural Network*” (*HNN*), que se abordan en la sección 4.2.6.

Las técnicas anteriores han sido utilizadas respectivamente en Pajares y Cruz (2006) y Pajares y col. (1998) en procesos de correspondencia estereoscópica de segmentos de bordes rectos, si bien bajo la estrategia propuesta en la rama derecha de la figura 4.2. Por tanto, de estas técnicas sólo se utiliza en este trabajo la parte conceptual relativa a la aplicación de los principios de similitud y proximidad determinados por la Gestalt (Koffka, 1935). Además, desde el planteamiento aquí propuesto, ambas técnicas, *FCM* y *HNN*, se enmarcan dentro de la filosofía de los *MRFs*, tal y como se expuso en el capítulo dos, sección 2.2.3.1.

En lo que sigue se detallan los procesos de aplicación de las distintas restricciones involucradas, según el esquema de la figura 4.2, siguiendo la estrategia de la rama izquierda. Así la restricción epipolar se describe en la sección 4.2.2; mientras que la de semejanza se expone en la sección 4.2.3; seguidamente se detalla la aplicación de la restricción de unicidad en la sección 4.2.4, tras la cual se construye el mapa de disparidad inicial, que se describe en la sección 4.2.5 y constituye el punto de partida para el siguiente proceso. Finalmente, en la sección 4.2.6, se describe el procedimiento de aplicación de la restricción de continuidad de la disparidad para filtrar el mapa inicial mediante los paradigmas *FCM* y *HNN*.

### 4.2.2 Epipolar: geometría del sistema

La figura 4.3 muestra la geometría del sistema de medición basado en visión estereoscópica (Herrera y col., 2009 $a,b,c,d$ ; Abraham y Förstner, 2005). El punto  $P$ , correspondiente a un hipotético objeto en la escena 3D con sus coordenadas del mundo tridimensional con respecto a los sistemas  $(X_1, Y_1, Z_1)$  y  $(X_2, Y_2, Z_2)$ , se proyecta en los puntos  $(x_{i1}, y_{i1})$  y  $(x_{i2}, y_{i2})$  en la imagen-1 (izquierda) y la imagen-2 (derecha) respectivamente en coordenadas del sistema de la imagen. Por otro lado,  $\alpha_1$  y  $\alpha_2$  son los ángulos de incidencia de los rayos de proyección desde  $P$ ;  $y_{12}$  representa la *línea base* que mide la distancia entre los ejes ópticos en lo que serían las dos posiciones de la cámara a lo largo de los respectivos ejes y de coordenadas;  $r$  es la distancia entre el punto proyectado en la imagen y el eje óptico correspondiente; finalmente  $R$  es el radio de la imagen, idéntico en ambas imágenes.

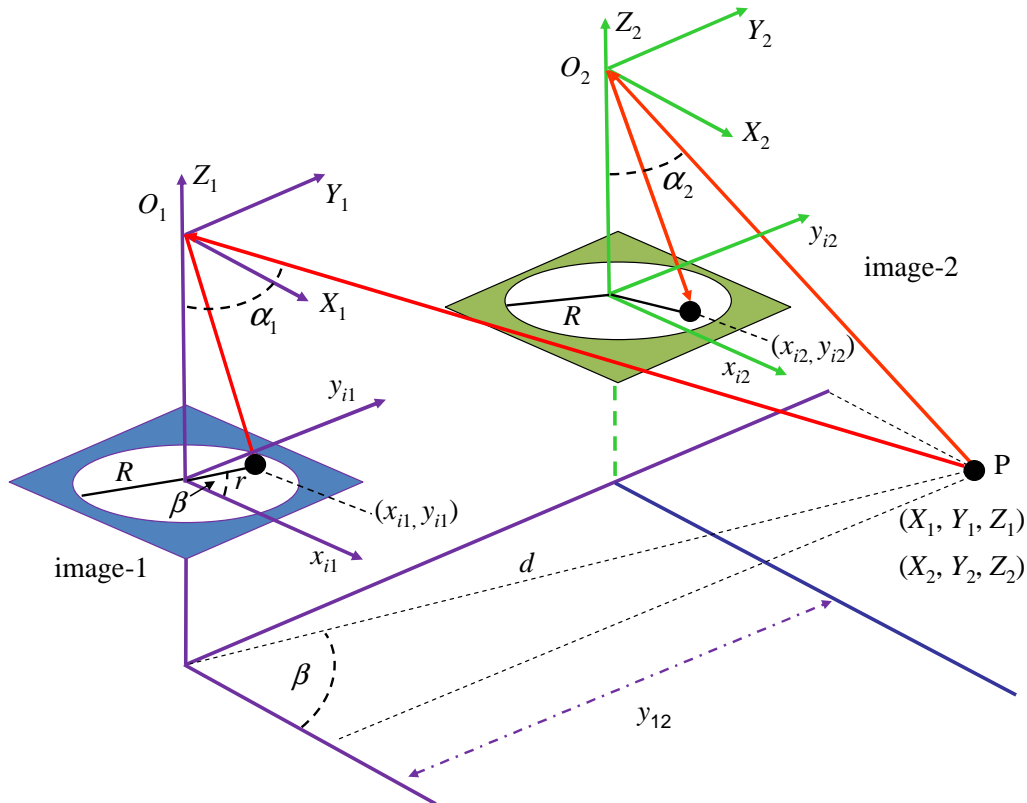


Figura 4.3 Proyecciones geométricas y relaciones para el sistema de visión estéreo basado en lentes de ojo de pez.



De acuerdo con Schwalbe (2005), pueden establecerse las siguientes relaciones geométricas:

$$r = \sqrt{x_{i1}^2 + y_{i1}^2}; \quad \alpha_1 = \frac{\pi r}{2R}; \quad \beta = \tan^{-1}(y_{i1}/x_{i1}) \quad (4.1)$$

El problema estriba en que las coordenadas del mundo 3D asociadas al mencionado punto  $P$ , esto es  $(X_1, Y_1, Z_1)$ , resultan desconocidas. Las mismas se pueden obtener a través de la ecuación (4.2) siempre y cuando la distancia  $d$  sea conocida.

$$X_1 = d \cos \beta; \quad Y_1 = d \sin \beta; \quad Z_1 = \sqrt{X_1^2 + Y_1^2} / \tan \alpha_1 \quad (4.2)$$

A partir de (4.2) es posible transformar las coordenadas del mundo en el sistema  $O_1X_1Y_1Z_1$  a las coordenadas del mundo en el sistema  $O_2X_2Y_2Z_2$  teniendo en cuenta la *línea base*,  $y_{12}$  como sigue:

$$X_2 = X_1; \quad Y_2 = Y_1 + y_{12}; \quad Z_2 = Z_1 \quad (4.3)$$

Suponiendo que la distorsión radial introducida por las lentes es despreciable, se pueden encontrar las coordenadas del punto  $P$  en el espacio 3D proyectadas sobre la imagen-2 a través de la ecuación (4.4) (Schwalbe, 2005).

$$x_{i2} = \frac{2R \arctan(\sqrt{X_2^2 + Y_2^2}/Z_2)}{\pi \sqrt{(Y_2/X_2)^2 + 1}}, \quad y_{i2} = \frac{2R \arctan(\sqrt{X_2^2 + Y_2^2}/Z_2)}{\pi \sqrt{(X_2/Y_2)^2 + 1}} \quad (4.4)$$

Debido a la geometría del sistema y la proyección basada en lentes de esta naturaleza, las líneas epipolares no son circunferencias concéntricas con respecto al centro de la imagen, lo que implica que este hecho debe ser considerado durante el proceso de correspondencia. La figura 4.4 muestra cuatro líneas epipolares, en el tercer cuadrante de la imagen derecha, que han sido generadas por los cuatro píxeles ubicados en las posiciones marcadas con los cuadrados rojos, que corresponden exactamente a sus posiciones equivalentes en la imagen izquierda. Estas líneas sirven de guía al proceso de correspondencia, de suerte que cuando se está tratando de buscar píxeles homólogos desde la imagen izquierda hacia la derecha, el recorrido de la línea epipolar se realiza en el sentido que marca el recorrido desde el Hemisferio Norte en la imagen derecha hacia el Sur, mientras que cuando la búsqueda de correspondencias se realiza

desde la imagen derecha hacia la izquierda, el recorrido de la línea epipolar se realiza desde el Hemisferio Sur hacia el Norte en la imagen izquierda.

Utilizando sólo una cámara en una única posición, digamos la  $O_1X_1Y_1Z_1$ , se obtendría solamente una imagen, de suerte que los puntos 3D pertenecientes a la línea  $\overline{O_1P}$  se proyectan todos sobre el punto único representado como  $(x_{i1}, y_{i1})$ . De este modo, con esa única imagen, no pueden obtenerse las coordenadas 3D del punto  $P$ , quedando clara por tanto la necesidad de las dos imágenes obtenidas desde posiciones diferentes. Cuando se intenta establecer la correspondencia del punto proyectado  $(x_{i1}, y_{i1})$  en la imagen-2 se sigue la línea epipolar, es decir, la proyección de  $\overline{O_1P}$  sobre la imagen-2. Esto, en términos operativos, es equivalente a variar el parámetro  $d$  en el espacio 3D. Así, dado el punto proyectado  $(x_{i1}, y_{i1})$  en la imagen-1 y siguiendo la línea epipolar, se obtiene una lista de  $m$  posibles píxeles que son correspondencias candidatas representadas cada uno por un punto de coordenadas  $(x_{i2}, y_{i2})$  en la imagen-2. De entre todas ellas, la considerada como la mejor correspondencia se asocia a una distancia  $d$  para el punto  $P$  en la escena 3D, que se calcula a partir de la geometría del sistema de visión estereoscópica. Por tanto, para cada  $d$  se obtiene un  $(x_{i2}, y_{i2})$  específico, de manera que cuando se establezca la correspondencia con  $(x_{i1}, y_{i1})$  por cualquiera de los procesos diseñados al respecto,  $d$  será la distancia a la que se encuentra el punto  $P$ .

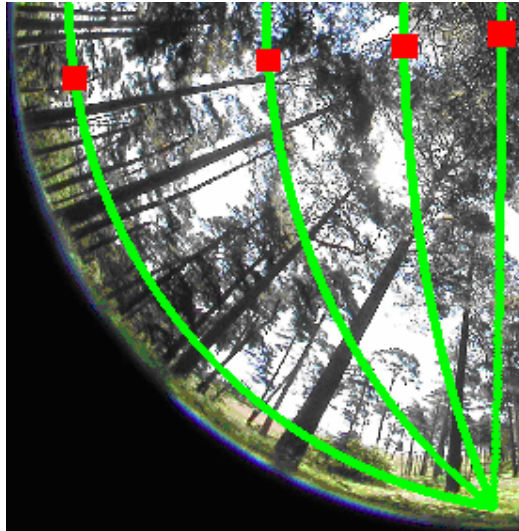


Figura 4.4 Líneas epipolares en la imagen derecha generadas a partir de las posiciones en la imagen izquierda marcadas con los cuadrados.

Si se toman medidas de distancias a los mismos puntos o estructuras específicas en los árboles tales como a la base y el final del tronco, la anchura del tronco medida a la misma altura del suelo u otros establecidos por los expertos, durante diferentes intervalos de tiempo, quizás años, es posible determinar la evolución del árbol y en consecuencia su estado de crecimiento. También el volumen de madera del mismo y naturalmente la densidad de árboles en la zona de exploración.

Naturalmente, lo anterior requiere que el sistema de visión estereoscópica esté ubicado en la misma posición geográfica en la escena 3D, para que el cómputo de las variables anteriormente mencionadas sea coherente en relación con las medidas tomadas a lo largo del tiempo. Esta ubicación es relativamente fácil, ya que los bosques están divididos en parcelas perfectamente identificadas geográficamente, así como los centros de cada una de ellas. El sistema se coloca en el centro de la parcela, orientado geográficamente siempre de igual manera, esto es, por ejemplo la posición de la cámara para obtener la imagen izquierda hacia el Sur geográfico y la posición de la cámara para la obtención de la imagen derecha hacia el Norte. Esta misma circunstancia hará que los árboles situados en diferentes zonas geográficas se proyectarán de diferente manera en las dos imágenes del par estereoscópico, lo que inducirá a una particularización especial a la hora de aplicar la restricción de unicidad, tal y como se describe en la sección 4.2.4. Todos estos detalles serán posteriormente ampliados en el capítulo cinco, sección 5.1.1.

### 4.2.3 Semejanza: atributos basados en el área y en las características

Cada píxel  $l$  en la imagen izquierda se caracteriza por sus atributos; cada uno de estos atributos se denota por  $A_l$ . Del mismo modo, cada candidato  $i$  en la lista de  $m$  candidatos en la imagen derecha se describe por atributos idénticos,  $A_i$ . Por tanto, se pueden calcular diferencias entre atributos del mismo tipo  $A$ , obteniendo una medida de semejanza para cada atributo como sigue:

$$s_{iA} = (1 + |A_l - A_i|)^{-1}; \quad i = 1, \dots, m \quad (4.5)$$

$s_{iA} \in [0, 1]$ ,  $s_{iA} = 0$  si la diferencia entre los atributos es suficientemente grande (mínima similitud), por otra parte si son iguales  $s_{iA} = 1$  (máxima similitud).

En el capítulo tres, sección 3.2.4, se realizó una descripción detallada de los seis atributos utilizados para describir cada característica, en relación al caso que nos ocupa, esto es, píxeles. En concreto son:

- a) coeficiente de correlación
- b) textura
- c) color en el espacio *RGB*
- d) módulo del gradiente
- e) ángulo del gradiente
- f) Laplaciana.

Siguiendo la terminología de la literatura especializada, los dos primeros son catalogados como basados en el área (“*area-based*”), y se obtienen sobre un entorno de vecindad de dimensión  $n \times n$ , en nuestro caso  $3 \times 3$ , alrededor de cada píxel por medio del coeficiente de correlación y la desviación estándar, computados respectivamente sobre dicha ventana. Los cuatro atributos restantes se consideran basados en las características (“*feature-based*”) (Lew y col. 1994). Puesto que el color involucra las tres componentes espectrales rojo-verde-azul (*RGB*), el valor absoluto en la ecuación (4.5) se extiende a las tres componentes como sigue:  $|A_l - A_i| = \sum_H |H_l - H_i|$ ,  $H = R, G, B$ .

Como ya se mencionó en el capítulo tres, ambos, color y módulo del gradiente se han combinado linealmente en Klaus y col. (2006) obteniendo resultados satisfactorios comparados con las imágenes estereoscópicas procedentes del repositorio correspondiente al banco de pruebas de Middlebury (Scharstein y Szeliski, 2002; Middlebury, 2010). Los coeficientes involucrados en la mencionada combinación lineal se calculan mediante pruebas de correspondencia fiables en una serie de experimentos llevados a cabo durante una etapa anterior.

Una vez identificados los atributos a utilizar, dado un píxel en la imagen izquierda y el conjunto de  $m$  candidatos en la derecha, a partir de ellos se obtienen las siguientes medidas de similitud para cada atributo  $A$ :  $s_{ia}$  (correlación),  $s_{ib}$  (color),  $s_{ic}$  (textura),  $s_{id}$  (módulo del gradiente),  $s_{ie}$  (ángulo del gradiente) y  $s_{if}$  (Laplaciana). Los identificadores en los subíndices identifican los atributos de acuerdo con las asignaciones anteriores. El mismo proceso se aplica en el caso de que el píxel de referencia proceda de la imagen derecha y los  $m$  candidatos de la izquierda.

#### 4.2.4 Unicidad especial: métodos basados en toma de decisiones

##### *a) Consideraciones derivadas de la geometría del sistema*

Llegados a este punto, antes de proceder a describir la aplicación de la restricción de unicidad, analizamos las consecuencias sobre dicha restricción derivadas de la geometría del sistema. Para ello y por comodidad, reproducimos de nuevo en la figura 4.5 el par de imágenes estereoscópicas mostradas en la figura 1.3. Sin pérdida de generalidad y teniendo en cuenta la disposición de las imágenes en las parcelas, tal y como se ha descrito en la sección 4.2.2, supongamos que la imagen izquierda se obtiene con la cámara orienta hacia el Sur geográfico y la derecha con la cámara orientada hacia el Norte, tal y como se describe en la sección 5.1.1. La figura 4.6 trata de reproducir de forma esquemática la situación real en la escena 3D. En dicha figura el círculo blanco representaría la zona del bosque bajo exploración, y sobre dicha zona se sitúa el sistema de medición con las ubicaciones marcadas como D e I, que identifican las posiciones donde se sitúa la cámara para obtener las respectivas imágenes derecha e izquierda. En esta disposición se ha tratado de mantener la ubicación de D hacia el Norte e I hacia el Sur. Volviendo sobre las imágenes de la figura 4.5, en la imagen de la izquierda denotada como (a) se aprecia una marca claramente visible en su parte inferior. Dicha marca también aparece en la imagen de la derecha (b), si bien mucho más reducida. Se trata de una varilla vertical situada sobre el soporte físico del sistema de visión estereoscópica, cuya orientación es hacia el Sur geográfico, es decir, en la posición I de la figura 4.6. De esta forma la zona del bosque situada hacia el Sur respecto del sistema estereoscópico se proyecta en sendas imágenes sobre el Hemisferio Sur (HS). Como quiera que la zona geográfica proyectada sobre este hemisferio se encuentra más próxima a la posición desde la que se obtiene la imagen izquierda que a la posición desde la que se obtiene la derecha, los troncos de los árboles situados en HS se proyectan en la imagen izquierda con un grosor mayor que sobre la imagen derecha. Por el contrario, la posición de la cámara desde la que se captura la imagen izquierda se encuentra más alejada de la zona geográfica orientada hacia el Norte geográfico, con respecto al sistema estereoscópico, que la posición de la cámara desde la que se obtiene la imagen derecha. Como en el caso anterior este hecho hace que los troncos del Hemisferio Norte (HN) de la escena 3D se proyecten sobre la imagen derecha con un

área mayor que las correspondientes proyecciones de los mismos troncos en la imagen izquierda del mismo hemisferio.

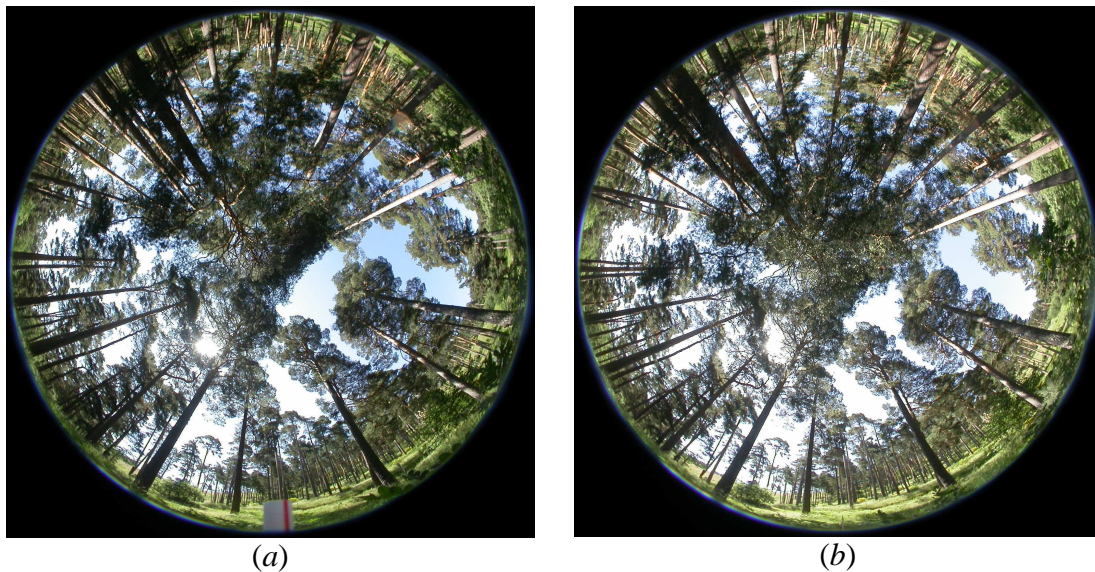


Figura 4.5 Par de imágenes estereoscópicas omnidireccionales: (a) izquierda, orientada hacia el Sur geográfico; (b) derecha, orientada hacia el Norte geográfico.

Las consecuencias inmediatas derivadas de esta circunstancia afectan directamente al proceso de correspondencia en lo que respecta a la restricción de unicidad. En efecto, desde el punto de vista del proceso de correspondencia es necesario realizar una serie de consideraciones, para lo que nos servimos del esquema mostrado en la figura 4.7. En la parte (a) de esta última figura, sobre el HN, aparece un tronco de árbol y el mismo tronco en la (b). La consecuencia de la proyección sobre este hemisferio es el mayor área que dicho tronco manifiesta en la imagen derecha. El caso contrario ocurre en el HS donde aparece simbolizada la proyección del mismo tronco 3D sobre ambas imágenes. En este caso puede observarse cómo el área es mayor en la imagen izquierda que en la derecha. Pues bien, las consideraciones anteriormente mencionadas a la hora de aplicar la restricción de unicidad se concretan como sigue:

- 1) Un píxel situado en el HN de la imagen izquierda puede y debe corresponderse con más de un píxel en el mismo hemisferio de la imagen derecha.
- 2) Un píxel situado en el HS de la imagen derecha puede y debe corresponderse con más de un píxel en el mismo hemisferio de la imagen izquierda.

- 3) En la zona de separación de ambos hemisferios, las correspondencias a nivel de píxel son generalmente uno a uno.

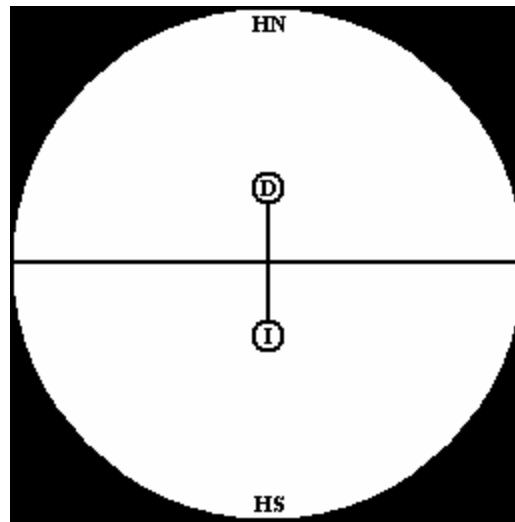


Figura 4.6 Orientación de las cámaras en la escena 3D

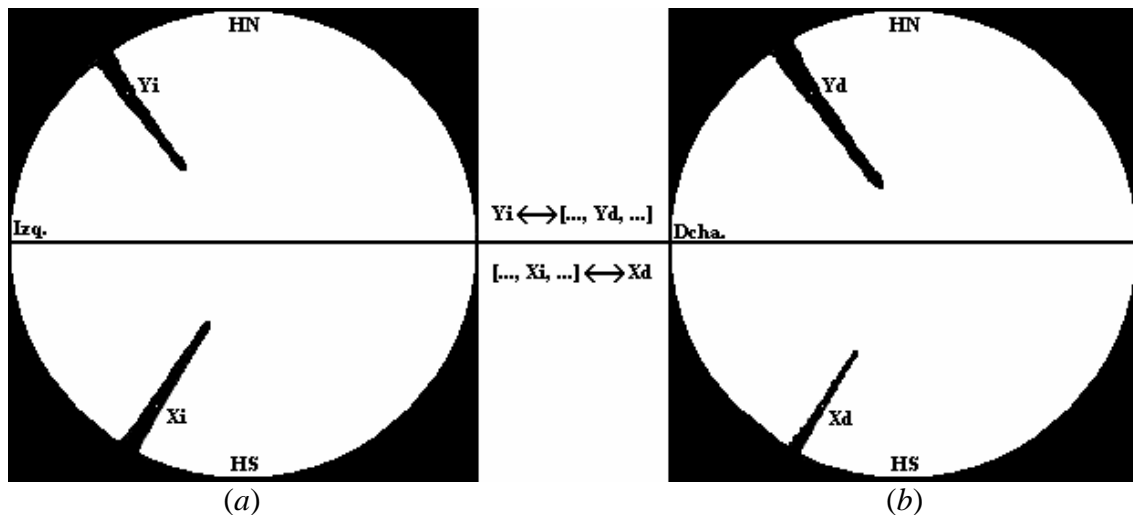


Figura 4.7 Esquema a considerar para la aplicación de la restricción de unicidad especial; (a) imagen izquierda; (b) imagen derecha

#### *b) Implicaciones en el proceso de toma de decisiones*

Basándose en las conclusiones presentadas en Klaus y col. (2006), la combinación de atributos aparece como un enfoque apropiado para mejorar los resultados de la correspondencia estereoscópica. Los métodos que se proponen a continuación tienen la capacidad de aplicar específicamente la combinación de atributos ya que en su origen están diseñados para combinación de clasificadores (Kuncheva, 2004). En el presente trabajo de investigación se realiza el ajuste apropiado para que estas técnicas de combinación de clasificadores puedan desempeñar el importante papel

que se les asigna en el proceso de correspondencia en visión estereoscópica. El ajuste consiste en utilizar los métodos basados en la teoría de toma de decisiones, para seleccionar uno o varios candidatos, según la situación de los píxeles en uno u otro hemisferio, teniendo en cuenta las tres consideraciones reseñadas previamente. La adaptación de estos mecanismos de decisión al caso que nos ocupa, nos sitúa en una posición inmejorable para aplicar conjuntamente las restricciones de similitud y unicidad de un modo apropiado.

Resumiendo lo expuesto hasta aquí, el proceso de correspondencia en general contempla los siguientes aspectos:

- 1) En primer lugar se establecen las posibles correspondencias entre la imagen izquierda y la derecha tomando como referencia la primera. Seguidamente, se determinan las correspondencias entre la imagen derecha y la izquierda con referencia en la derecha. Con este mecanismo se trata de considerar los aspectos derivados de la geometría y posicionamiento del sistema en la línea anteriormente expuesta en el apartado *a*). Además, con ello se consideran algunas iniciativas existentes en la literatura sobre correspondencia estereoscópica, en el sentido de que si una determinada característica en una imagen elige como preferencia a una cierta característica de la otra imagen, esta última también debería preferir a la primera para un correcto emparejamiento (Ogale y Aloimonos, 2005; Elias, 2007).
- 2) Los mecanismos basados en la toma de decisión son los encargados de aplicar la restricción de unicidad de forma extendida, estableciendo las correspondencias bajo la consideración de que un píxel puede tener más de una correspondencia según su ubicación en cada hemisferio de las imágenes y según que el proceso de correspondencia se establezca de izquierda a derecha o viceversa. En cualquier caso, cuando se tolere la existencia de más de un candidato, es necesario considerar el hecho de que éstos se encuentran localizados en posiciones espaciales consecutivas de la imagen, ya que la mayor extensión en el área de los troncos en la imagen se realiza sobre la base de que los puntos del espacio, que originan las imágenes, se ubican en posiciones consecutivas.



- 3) Finalmente en el mecanismo de decisión se incluye la posibilidad de que no exista correspondencia para un píxel dado, debido a las posibles oclusiones, que pueden aparecer como consecuencia de que determinados troncos ocultan a otros en una de las imágenes pero no lo hacen en la otra del par estereoscópico.

En las siguientes cuatro subsecciones se describen los cuatro mecanismos propuestos en el presente trabajo de investigación para establecer las correspondencias entre píxeles por aplicación de la restricción adaptada de unicidad sobre la base de la restricción de similitud.

#### 4.2.4.1 Integral Fuzzy de Choquet (CFI)

Bajo el paradigma *CFI* (Meyer y Roubens, 2006; Kuncheva, 2004) el objetivo consiste en hacer corresponder cada píxel  $l$  en la imagen izquierda con el mejor de los candidatos posibles en la imagen derecha si  $l$  se encuentra en el HS; mientras que si  $l$  se sitúa en el HN, se tolera más de un candidato para  $l$  siempre y cuando se cumplan las propiedades que se detallan posteriormente. De igual modo para el caso de una correspondencia cuya referencia es la imagen derecha, donde ahora el proceso trata de emparejar un píxel  $m$  en la imagen derecha con un único píxel en la izquierda si  $m$  se sitúa en el HN de esta imagen o con varios si se ubica en el HS. Se trata pues de una decisión basada en el paradigma *CFI*, el cual permite combinar las similitudes a nivel de píxel (característica), que se calculan mediante la ecuación (4.5).

El método *CFI* requiere el cómputo de la relevancia asignada a cada atributo, a partir de la cual se pueden obtener las denominadas *densidades fuzzy*. Esto se resuelve calculando una medida conocida como  $\lambda$ -fuzzy. Para ello se requiere disponer de una serie de datos previos, es decir, un conjunto de muestras formadas por pares de correspondencias seleccionadas previamente por algún procedimiento. En nuestro caso estas correspondencias se establecen manualmente, de forma similar a como se realiza en Kuncheva (2004).

El proceso comienza calculando un conjunto de seis valores *fuzzy*,  $g^a, g^b, g^c, g^d, g^e, g^f$ , cada uno de los cuales representa la relevancia individual del atributo asociado. Los atributos son los seis descritos anteriores, es decir,  $\Omega \equiv \{a, b, c, d, e, f\}$  asociados

respectivamente a la correlación, color, textura, módulo del gradiente, ángulo del gradiente y Laplaciana. Por tanto, según esta nomenclatura,  $g^a$  se refiere a la relevancia del atributo correlación,  $g^b$  del atributo color y así sucesivamente.

Las relevancias individuales para cada atributo se calculan a partir de los datos seleccionados. La forma de obtenerlos varía dependiendo de la aplicación. En nuestro caso, el proceso de asignación se realiza mediante el procedimiento descrito en el capítulo cinco, sección 5.4.1.1.

El valor de  $\lambda$  necesario para calcular las  $g$  densidades *fuzzy* se obtiene como la única raíz real mayor que  $-1$  del polinomio siguiente:

$$\lambda + 1 = \prod_{j \in \Omega} (1 + \lambda g^j) \quad (4.6)$$

A partir de las relevancias individuales  $g^a, \dots, g^f$  y una vez calculada  $\lambda$ , el procedimiento basado en la integral *fuzzy* adaptado a nuestro problema de correspondencia estereoscópica píxel a píxel queda establecido como sigue:

1. Dado un píxel  $l$  bien en la imagen izquierda o en la derecha, a partir de la ecuación (4.5) se calculan las similitudes entre  $l$  y cada candidato  $i$  en la otra imagen del par estereoscópico, siendo  $i = 1, \dots, m$ , obteniendo un vector columna de la forma:  $[s_{ia}, s_{ib}, s_{ic}, s_{id}, s_{ie}, s_{if}]^T$ , donde cada una de las componentes representa la similitud existente entre el píxel  $l$  y el  $i$  según el atributo  $A$  en el conjunto  $\Omega \equiv \{a, b, c, d, e, f\}$ . Sin pérdida de generalidad se supone que  $s_{ia}$  es el valor de similitud más alto y  $s_{if}$  el más bajo. Dicho vector se ordena bajo este criterio, es decir, según la relación siguiente:  $s_{ia} > s_{ib} > s_{ic} > s_{id} > s_{ie} > s_{if}$ .
2. Basándonos en el orden de similitudes establecido previamente, se ordenan también los correspondientes valores *fuzzy*, asignándose a cada atributo su correspondiente valor *fuzzy*, es decir, el orden queda establecido como sigue:  $g^a, g^b, g^c, g^d, g^e, g^f$ . A continuación se obtiene el primer valor de densidad *fuzzy* tal que,  $g(a) = g^a$ .
3. A partir del valor anterior se calculan el resto de densidades *fuzzy* según el procedimiento recursivo descrito en la ecuación (4.7).

$$\begin{aligned}
g(b) &= g^b + g(a) + \lambda g^b g(a) \\
g(c) &= g^c + g(b) + \lambda g^c g(b) \\
&\dots\dots\dots \\
g(f) &= g^f + g(e) + \lambda g^f g(e)
\end{aligned} \tag{4.7}$$

4. Para cada candidato  $i$  se calcula el grado de correspondencia entre este  $i$  y el píxel  $l$  que originó la lista de candidatos como sigue:

$$S_i(l) = s_{ia} + \sum_{\substack{h=b \\ \Omega}}^f \left[ s_{i(h-1)} - s_{ih} \right] g(h-1) \tag{4.8}$$

donde los términos con  $s$  minúscula se refieren a similitudes obtenidas a partir de la ecuación (4.5) sobre el conjunto de sus atributos  $\Omega$ . El sumatorio en la ecuación (4.8) representa el término integral basado en las diferencias de semejanza entre dos valores de similitud consecutivos según el orden decreciente establecido previamente sobre dichos valores, que representarían las bases de los hipotéticos rectángulos de integración con alturas  $g(h-1)$ , con  $h$  variando desde  $b$  hasta  $f$  en el conjunto ordenado  $\Omega$ .

5. La decisión consta de un primer paso donde se ha de tener en cuenta la ubicación del píxel  $i$  en los hemisferios HS o HN y si la correspondencia se está estableciendo de izquierda a derecha o de derecha a izquierda, para decidir si se selecciona un único candidato o varios, según lo expuesto anteriormente. En el caso de un único candidato, la decisión sobre la mejor correspondencia se toma seleccionando el máximo valor del grado de correspondencia  $S_i(l)$  entre todos los candidatos  $i$  posibles, siempre y cuando el grado de correspondencia supere un valor umbral prefijado. En el caso de tener la opción de varios candidatos, se seleccionan todos aquellos que superen el valor umbral prefijado teniendo en cuenta que estos varios candidatos deben situarse contiguos espacialmente para garantizar que proceden de la misma entidad física, si bien bajo una proyección de la característica expandida, dependiendo de los hemisferios en las imágenes, tal y como se ha explicado previamente. El umbral anteriormente mencionado, tiene como objetivo garantizar que las correspondencias sean lo suficientemente fiables, con el fin de evitar falsas correspondencias procedentes de oclusiones u

otros fenómenos espurios. En este trabajo se ha fijado al valor de 0.7, al considerar que es un umbral relativamente exigente. Esto podría dar lugar a situaciones en las que un píxel no encuentre correspondencia en uno de los dos sentidos de búsqueda o en ninguno, en cuyo caso se le asigna un valor de disparidad nulo, que en realidad representa un valor desconocido.

6. El segundo paso de decisión consiste en la eliminación de las posibles ambigüedades que todavía puedan persistir tras la aplicación del punto anterior. Para ello se consideran las listas de candidatos provenientes de la correspondencia izquierda a derecha y derecha a izquierda, y se eligen finalmente los emparejamientos que coinciden en ambos sentidos de la correspondencia, eso sí, teniendo en cuenta los hechos singulares de ubicación de los píxeles según los hemisferios de las imágenes. Como se verá posteriormente en la sección 4.2.5, para la elaboración del mapa de disparidad se tomará como referencia la imagen izquierda, si bien se podría haber tomado la derecha sin ningún tipo de influencia ni a favor ni en contra del proceso global de correspondencia. Por tanto, dado un píxel en la imagen izquierda, se buscan para él todas las parejas en las que interviene, tanto en el sentido izquierda-derecha como derecha-izquierda. En relación a las posibles parejas que se puedan formar caben todavía cuatro posibilidades:

- a) Sólo ha quedado un único emparejamiento, siendo éste el que se elige como la correspondencia entre ambos píxeles del par.
- b) Todavía persiste ambigüedad en el sentido de que hay más de una pareja en la que está involucrado el píxel de la imagen izquierda, en cuyo caso se elige como único emparejamiento la que haya obtenido el mayor grado de correspondencia. En caso de que haya más de un par con el mismo grado se eligen todos ellos y la disparidad se obtiene como la media de los valores de disparidad entre los pares ambiguos.
- c) Un píxel de la izquierda sólo aparece como miembro de las parejas en relación a una única imagen, no encontrando pareja en la otra. Por ejemplo, un píxel de la imagen izquierda aparece emparejado con píxeles de la derecha, pero ese mismo píxel nunca aparece como miembro de

una pareja con origen en la imagen derecha. En estos casos se elige la pareja con el máximo grado de correspondencia de entre las que aparece.

- d) Finalmente, si un píxel de la izquierda no encuentra homólogo bien porque él mismo no se empareja con ninguno de la derecha o porque no es seleccionado por otros en la imagen derecha, entonces a este píxel se le asigna un valor de disparidad desconocido, que en nuestro caso es el valor nulo, ya que probablemente se trate de algún píxel procedente de alguna oclusión, tal y como se ha indicado anteriormente.

Tras la aplicación de este paso, cada píxel de la imagen izquierda termina emparejándose con un único píxel en la derecha.

#### 4.2.4.2 Integral Fuzzy de Sugeno (SFI)

El procedimiento basado en la Integral *Fuzzy* de Sugeno (*SFI*) es muy similar al *CFI*, coincidiendo exactamente en los tres primeros pasos y diferenciándose en el 4, 5 y 6, que quedan ahora reformulados como sigue,

4. Para cada candidato  $i$  se calcula el grado de correspondencia entre este  $i$  y el píxel  $l$  que originó la lista de candidatos como sigue:

$$\mu_i(l) = \max_{h \in \Omega} \{ \min \{ s_{ih}, g(h) \} \} \quad (4.9)$$

5. El proceso de decisión basado en el *SFI* sigue exactamente los mismos dos pasos descritos en los puntos cinco y seis de la sección anterior, si bien tomando como grados de correspondencia los obtenidos por medio de la ecuación (4.9). Se elige igualmente, el mismo valor de umbral como referencia para garantizar un alto grado de correspondencia.

#### 4.2.4.3 Teoría de Dempster-Shafer (DES)

Como en los casos anteriores relativos al *CFI* y *SFI* el problema consiste en establecer la correspondencia entre cada píxel  $l$  en la imagen izquierda con el mejor de los  $m$  candidatos posibles (unicidad) en la derecha. En este caso, la decisión se basa en la teoría proporcionada por el método *DES*. A continuación describimos brevemente este enfoque.

La teoría de Dempster-Shafer debe su nombre a los trabajos de ambos autores recogidos en Dempster (1968) y Shafer (1976). Seguidamente se detalla la aplicación de este método a nuestro problema de correspondencia siguiendo las directrices expuestas en Kuncheva (2004):

1. En nuestro problema de correspondencia estereoscópica un píxel  $l$  de la imagen izquierda puede corresponderse con otro  $i$  de la imagen derecha de forma correcta o incorrecta. Basándonos en este hecho, identificamos dos clases, la clase de las correspondencias positivas y la clase de las correspondencias negativas,  $w_1$  (positiva) y  $w_2$  (negativa), que en realidad representan emparejamientos correctos e incorrectos respectivamente. A partir de un conjunto de muestras de ambas clases, calculamos las similitudes de las correspondencias pertenecientes a cada una de las clases según la ecuación (4.5), construyendo a continuación un vector de medias de seis dimensiones, donde sus componentes son las medias de los valores de sus similitudes, es decir,  $\bar{\mathbf{v}}_j = [\bar{s}_{ja}, \bar{s}_{jb}, \bar{s}_{jc}, \bar{s}_{jd}, \bar{s}_{je}, \bar{s}_{jf}]^T$ ; con  $\bar{\mathbf{v}}_1$  y  $\bar{\mathbf{v}}_2$  siendo las medias para las clases  $w_1$  y  $w_2$  respectivamente. Esto se lleva a cabo durante una fase previa, equivalente a la de entrenamiento en problemas de clasificación y por tanto, también equivalente a las correspondientes fases para determinar las relevancias en las estrategias *CFI* y *SFI* vistas anteriormente o los pesos asociados en la *MCDMF*, que se verá posteriormente.
2. A partir de un candidato  $i$  en la lista de  $m$  candidatos para  $l$ , calculamos el vector de seis dimensiones  $\mathbf{x}_i$ , donde sus componentes son los valores de similitud obtenidos de acuerdo con (4.5) entre  $i$  y  $l$ , es decir  $\mathbf{x}_i = [s_{ia}, s_{ib}, s_{ic}, s_{id}, s_{ie}, s_{if}]^T$ . Seguidamente calculamos la proximidad  $\Phi$  entre cada componente en  $\mathbf{x}_i$  y cada componente en  $\bar{\mathbf{v}}_j$  basada en la norma Euclídea  $\|\cdot\|$ , ecuación (4.10).

$$\Phi_{jA}(\mathbf{x}_i) = \frac{\left(1 + \|s_{iA} - \bar{s}_{jA}\|^2\right)^{-1}}{\sum_{k=1}^2 \left(1 + \|s_{iA} - \bar{s}_{kA}\|^2\right)^{-1}} \text{ donde } A \in \Omega \quad (4.10)$$

3. Para cada clase  $w_j$  y para cada candidato  $i$ , calculamos lo que denominamos el *grado de confianza* a través de la ecuación (4.11).

$$b_j^i(A) = \frac{\Phi_{jA}(\mathbf{x}_i) \prod_{k \neq j} (1 - \Phi_{kA}(\mathbf{x}_i))}{1 - \Phi_{jA}(\mathbf{x}_i) \left[ 1 - \prod_{k \neq j} (1 - \Phi_{kA}(\mathbf{x}_i)) \right]}; \quad j = 1, 2 \quad (4.11)$$

4. El grado de pertenencia que cada candidato  $i$ , representado por  $\mathbf{x}_i$ , recibe para cada clase  $w_j$  teniendo en cuenta que su correspondencia es  $l$  viene dado por:

$$\mu_j(\mathbf{x}_i) = \prod_{A \in \Omega} b_j^i(A) \quad (4.12)$$

5. Llegados a este punto, el proceso de decisión sigue los mismos dos pasos enumerados como cinco y seis, descritos en la sección 4.2.4.1, donde ahora las correspondencias se establecen en función de los grados de pertenencia definidos en la ecuación (4.12) bajo el criterio del máximo, tomando como referencia el mismo valor umbral 0.7.

#### 4.2.4.4 Toma de Decisiones Multicriterio Fuzzy (MCDMF)

De nuevo, como en los tres casos anteriores, el objetivo para aplicar esta técnica consiste en hacer corresponder cada píxel  $l$  en la imagen izquierda con el mejor de los  $m$  candidatos posibles (unicidad) de la imagen derecha, representados por  $i$ . El mecanismo utilizado se fundamenta en el paradigma *MCDMF* (Chen, 2000; Wang y Fenton, 2006).

Para describir el proceso de decisión, necesitamos previamente definir los siguientes conceptos:

a) un número triangular fuzzy  $u$  como una terna  $(u_1, u_2, u_3)$

b) la distancia entre dos números triangulares fuzzy  $u$  y  $z$  como:

$$d(u, z) = \left\{ \left[ (u_1 - z_1)^2 + (u_2 - z_2)^2 + (u_3 - z_3)^2 \right] / 3 \right\}^{1/2} \quad (4.13)$$

Teniendo en cuenta el conjunto de las seis medidas de similitud introducidas a partir de la ecuación (4.5), las mismas se pueden distribuir en dos grupos que identificamos como  $C_1$  y  $C_2$ ; cada uno de dichos grupos conteniendo tres valores, a saber:  $C_1 = \{c_{11}, c_{12}, c_{13}\}$  y  $C_2 = \{c_{21}, c_{22}, c_{23}\}$ . De esta forma cada grupo define un criterio de similitud, y por tanto, se tienen disponibles dos criterios  $C_j$  ( $j = 1, 2$ ) para

tomar la decisión, según la adaptación de este enfoque a nuestro problema de correspondencia estereoscópica.

Bajo el paradigma *MCDMF* y asumiendo que existen  $m$  posibles candidatos, la decisión sobre el mejor de ellos puede formularse como la elección de la mejor alternativa  $M_i$  ( $i = 1, \dots, m$ ); donde ahora cada alternativa representa uno de esos posibles candidatos. El problema de la toma de decisión bajo el *MCDMF* puede expresarse en forma de matriz de la siguiente manera:

$$X = [x_{ij}]_{m \times n}; \quad W = [w_j]_{1 \times n}; \quad i = 1, \dots, m; \quad j = 1, \dots, n. \quad (4.14)$$

En la ecuación anterior,  $X$  es la matriz de decisión donde  $x_{ij}$  representa la valoración de la alternativa  $M_i$  frente al criterio  $C_j$ ;  $w_j$  es el peso asignado al criterio  $C_j$ , que representa de alguna manera la fuerza o importancia de dicho criterio a la hora de tomar la decisión.

Pues bien, con la formulación anterior y orientándonos hacia el problema de la correspondencia estereoscópica, se consideran las medidas de similitud como números triangulares *fuzzy* tales como  $x_{i1} = (s_{ia}, s_{ib}, s_{ic})$  y  $x_{i2} = (s_{id}, s_{ie}, s_{if})$ . Los pesos asociados a cada medida de similitud son respectivamente  $w_1 = (w_a, w_b, w_c)$  y  $w_2 = (w_d, w_e, w_f)$ , estimados de acuerdo a su relevancia o importancia relativa, como se describe más tarde en el capítulo cinco, sección 5.4.4.1. Sin pérdida de generalidad, los valores  $x_{i1}, x_{i2}$  se ordenan de modo que  $s_{ia} \leq s_{ib} \leq s_{ic}$  y también para  $s_{id} \leq s_{ie} \leq s_{if}$ . Este convenio se aplica exactamente igual sobre cada uno de los  $m$  candidatos disponibles. A partir de aquí, estamos en condiciones de obtener la matriz de decisión *fuzzy* normalizada como sigue:

$$R = [r_{ij}]_{m \times n}; \quad r_{i1} = \left( \frac{s_{ia}}{s_{ic}^*}, \frac{s_{ib}}{s_{ic}^*}, \frac{s_{ic}}{s_{ic}^*} \right); \quad r_{i2} = \left( \frac{s_{id}}{s_{if}^*}, \frac{s_{ie}}{s_{if}^*}, \frac{s_{if}}{s_{if}^*} \right) \quad (4.15)$$

donde  $s_{ic}^* = \max_i \{s_{ic}\}$  y  $s_{if}^* = \max_i \{s_{if}\}$ . Esto preserva la propiedad de que los rangos de los números triangulares *fuzzy* normalizados pertenecen al intervalo  $[0, +1]$ . Teniendo en cuenta la importancia asignada a cada criterio, se puede construir la matriz de decisión ponderada *fuzzy* normalizada según la ecuación (4.16).



$$V = [v_{ij}]_{m \times n} \quad \text{donde} \quad v_{i1} = r_{i1} w_{j1}; \quad v_{i2} = r_{i2} w_{j2} \quad (4.16)$$

En consecuencia, de  $V$  se conoce que los elementos  $v_{ij}, \forall ij$  son números triangulares *fuzzy* positivos normalizados variando en el intervalo cerrado  $[0, +1]$ . Por tanto se puede definir la solución ideal positiva *fuzzy* como  $p^+ = (1, 1, 1)$ , y la solución ideal negativa *fuzzy* como  $p^- = (0, 0, 0)$ . Las distancias a cada alternativa pueden calcularse de la siguiente manera:

$$d_i^+ = d(v_{i1}, p^+) + d(v_{i2}, p^+) \quad \text{y} \quad d_i^- = d(v_{i1}, p^-) + d(v_{i2}, p^-) \quad (4.17)$$

donde  $d(\cdot, \cdot)$  es la distancia medida entre dos números triangulares *fuzzy*, definida en la ecuación (4.13). De acuerdo con Chen (2000), el coeficiente de proximidad o “*closeness coefficient*” se define para determinar el orden de clasificación según las alternativas, una vez que ambas  $d_i^+$  y  $d_i^-$  han sido calculadas para cada alternativa. Este coeficiente resulta ser finalmente:

$$CC_i = d_i^- / (d_i^+ + d_i^-) \quad (4.18)$$

Obviamente, una alternativa  $M_i$  está más próxima a la solución ideal positiva *fuzzy* y más lejos de la solución ideal negativa *fuzzy* a medida que  $CC_i$  se aproxima a +1. Así, dado el píxel  $l$  en la imagen izquierda, su mejor correspondencia  $i$  de la lista de píxeles posibles en la imagen derecha, es aquella con el máximo valor para  $CC_i$  y lo mismo para el caso en el que el píxel  $l$  provenga de la imagen derecha y sus candidatos  $i$ , se sitúen en la izquierda. Por las mismas razones que las expuestas en el caso de las estrategias *CFI* y *SFI*, se aplican los mismos criterios que los descritos en los puntos cinco y seis de la sección 4.2.4.1, es decir, se buscan correspondencias en el doble sentido izquierda-derecha y se consideran una o varias correspondencias según la ubicación de los píxeles en los hemisferios de las imágenes. Igualmente, para garantizar las correspondencias fiables, se establece como umbral de referencia el mismo valor 0.7, fijado previamente para el resto de métodos anteriores.

### 4.2.5 Mapa de disparidad inicial

Por aplicación del proceso de decisión descrito en la sección 4.2.4.1 y por extensión en las tres siguientes subsecciones relativas a los procesos de decisión *SFI*, *DES* y *MCDMF*, para cada píxel  $l \equiv (x_l, y_l)$  en la imagen izquierda se dispone de su homólogo en la imagen derecha  $i \equiv (x_i, y_i)$ , cuyo emparejamiento se ha obtenido como el mejor de los posibles por cualquiera de los métodos anteriormente descritos.

Por tanto, de  $l$  e  $i$  se conocen sus posiciones en coordenadas Cartesianas, que son transformadas a coordenadas polares tomando el centro de la imagen como origen del sistema de referencia polar; así ambos píxeles  $l$  y  $i$  poseen ángulos  $\theta_l$  y  $\theta_i$  respectivamente en coordenadas polares. A partir de estos ángulos se construye un mapa con las mismas posiciones de píxeles que las de la imagen original izquierda, es decir,  $q = M \times N$ , donde obviamente cada posición representa un píxel. Dada la posición del píxel  $l \equiv (x_l, y_l)$ , se carga esa posición del mapa con el valor  $\Delta\theta_l = |\theta_l - \theta_i|$  que es exactamente la diferencia relativa en grados de ángulo sexagesimal entre los píxeles  $l$  e  $i$  y por tanto el valor de la disparidad buscada para el píxel  $l$ . Este proceso se lleva a cabo para todos los píxeles de la imagen original izquierda, excepto para aquellos que fueron clasificados como pertenecientes a las tres texturas consideradas fuera de interés, es decir, cielo, hierba en el suelo y hojas de los pinos, según el proceso de segmentación descrito en el capítulo tres y por tanto excluidas del proceso de correspondencia. Para los píxeles asociados a estas texturas, se les asigna un valor de disparidad nulo, coincidiendo, en este caso, con aquellos píxeles que se identificaron durante el proceso de decisión como posibles oclusiones. Los valores en el mapa de disparidad varían en el intervalo  $[0, \theta_{\max}]$ , donde  $\theta_{\max}$  se fija a 8.0 en nuestra propuesta, dado que resulta ser el máximo valor de disparidad observado en todas las imágenes estereoscópicas disponibles de este tipo.

De esta forma, el mapa de disparidad así obtenido es un mapa de disparidad inicial, que puede todavía mejorarse con el fin de homogeneizar la distribución de disparidades, en particular en las zonas de los troncos, que es donde realmente interesa. La propuesta que se formula en este trabajo de investigación, para llevar a cabo dicha homogeneización, se basa en los enfoques *FCM* y *HNN*, que se describen en la

siguiente sección, en base a la aplicación de la restricción de continuidad de la disparidad según el diseño planteado en la sección 4.2.1.

#### 4.2.6 Continuidad de la disparidad: filtrado del mapa de disparidad

Obtenido el mapa de disparidad inicial, de acuerdo con el proceso descrito previamente, el objetivo ahora se centra en la aplicación de la restricción de continuidad de la disparidad para la mejora de ese mapa inicial.

Cualquiera de los dos enfoques propuestos puede contemplarse desde el punto de vista de la computación neuronal, dentro de la capacidad sensorial por la cual una red percibe los objetos de la escena. Para tratar con escenas del mundo real es necesario establecer algún criterio que agrupe de alguna manera los elementos presentes en la escena. En el trabajo de Wang (2005) se estudia exhaustivamente una lista de los principios más importantes relativos a dicha agrupación desde el punto de vista de los problemas relacionados con la Visión por Computador. Éstos, a su vez, están inspirados en los principios de la Gestalt (Koffka, 1935). Desde el punto de vista de nuestra propuesta, se consideran los tres principios siguientes: *proximidad*, píxeles etiquetados que se encuentran próximos en el espacio tienden a agruparse; *semejanza*, píxeles etiquetados con valores similares tienden a agruparse; *conexión*, píxeles etiquetados que se encuentran dentro de la misma región espacial e interconectados tienden a agruparse. Este planteamiento sigue la filosofía de los *MRFs*, tal y como se expuso en la sección 4.2.1.

La aplicación de estos principios permite definir un espacio de vecindad alrededor de un píxel que constituye un elemento central a la hora de aplicar tanto el paradigma de los *FCMs* como de las *HNNs*, que son las dos estrategias propuestas en este trabajo.

A partir del mapa de disparidad inicial disponible en este momento, construimos una red de nodos, cuya topología se establece atendiendo a la distribución espacial del mapa de disparidad. De este modo, cada nodo en la red se localiza en la misma posición que los elementos en el mapa, es decir, en la misma posición que el píxel correspondiente en la imagen izquierda. Por tanto un nodo se identifica con el píxel  $l$ , y en consecuencia, el número de nodos en la red es  $q = M \times N$ .

En resumen, se dispone de una red de  $q$  nodos ubicados espacialmente en las mismas posiciones de los píxeles de la imagen izquierda. Para mantener la nomenclatura de la literatura especializada, a partir de ahora nos referimos a los nodos en la red con los símbolos  $i$  o  $k$  sin que ello suponga una confusión respecto de la nomenclatura genérica anteriormente utilizada para designar a los píxeles.

Un nodo  $i$  en la red se inicializa con el valor de disparidad obtenido a partir del mapa de disparidad en la misma posición, es decir,  $\Delta\theta_i$ , pero en lugar de utilizar el rango  $[0, \theta_{\max}]$  como valores posibles, éstos se transforman linealmente a los nuevos valores de disparidad en el rango  $[-1, +1]$ . Además, por simplicidad  $\Delta\theta_i$  se renombra como  $D_i$ , identificando a éste como el valor del estado inicial del nodo  $i$  por una cuestión relacionada con la dinámica de la red.

Además, entre dos nodos  $i$  y  $k$  existe una relación determinada por lo que se conoce como *peso causal*, tanto en los *FCMs* como en las *HNNs*, estos valores determinan la influencia ejercida por el nodo  $k$  sobre  $i$ , incrementando o disminuyendo el valor del estado del nodo  $i$ , dependiendo de la causalidad de los valores positivos o negativos, respectivamente. Si el peso de conexión es nulo, indica ausencia de causalidad.

Tanto el *FCM* como la *HNN* tienen como objetivo la modificación de los estados del nodo  $i$  en función de las influencias ejercidas por los nodos vecinos según el valor de sus estados y los pesos de conexión. Estas influencias se conocen como *influencias contextuales*. En este trabajo los pesos de conexión se obtienen a través de lo que denominamos *coeficientes de consistencia*, tanto en los *FCMs* como en las *HNNs*. Es evidente que la modificación de los estados de los nodos conlleva implícitamente la modificación de la disparidad asociada a cada nodo en cuestión. En conjunto, en ambos procesos se pretende conseguir estados de los nodos estables, de suerte que a medida que avanza el proceso iterativo, tanto en *FCM* como en *HNN*, los estados de los nodos, y por tanto las disparidades, tienden hacia esa deseada estabilización. Es en este momento cuando se dice que la red en su conjunto ha adquirido un estado global estable.

En la sección 4.2.6.1a se proporcionan los detalles específicos acerca de la topología de los *FCMs* y su modo de funcionamiento en la sección 4.2.6.1b. En la

sección 4.2.6.2a se describe la topología del *HNN* y en la sección 4.2.6.2b su mecanismo de optimización para la actualización de los nodos.

#### 4.2.6.1 Mapas Cognitivos Fuzzy (FCMs)

##### a) Proceso de actualización de los estados de los nodos

Desde el punto de vista de los *FCMs*, los nodos se consideran conceptos, con sus correspondientes relaciones causales entre ellos (Tsardias y Margaritis, 1997, 1999; Kosko, 1986, 1992; Miao y Liu, 2000). Los pesos causales  $w_{ik}$  representan los valores asignados a la relación causal existente entre los nodos  $i$  y  $k$  en la red, tomando valores en el intervalo causal *fuzzy*  $[-1, +1]$ . En los *FCMs* no se permite la retroalimentación de un nodo sobre sí mismo (Kosko, 1986, 1992), por consiguiente,  $w_{ii} = 0$ .

Bajo el enfoque de los *FCMs*, que es un proceso iterativo, los estados iniciales en la iteración  $t = 0$ ,  $D_i(t=0)$ , se actualizan basándose en la influencias causales en cada iteración  $t$ . Cada nodo se activa positiva o negativamente a un cierto nivel de activación. De acuerdo con Tsardias y Margaritis (1997, 1999), dicho nivel de activación en la iteración  $t + 1$  se calcula como:

$$D_i(t+1) = f(D_i(t), A_i) = f\left(D_i(t) + \sum_{k=1}^q w_{ik}(t) D_k(t)\right) - d_i D_i(t) \quad (4.19)$$

donde  $D_i(t)$ , definido previamente, representa la activación del nodo  $i$  en la iteración  $t$  en la red. Cada peso causal  $w_{ik}$  se define como un coeficiente de regularización que representa la influencia mutua ejercida por los  $k$  vecinos sobre  $i$ .  $A_i$  representa la suma de las influencias ponderadas que recibe el nodo  $i$  en la iteración  $t$  del resto de nodos. El término  $d_i \in [0, 1]$  es el factor de decadencia de la neurona  $i$ . Este factor determina la fracción del actual nivel de activación que se restará del nuevo nivel de activación como resultado de la intención natural del nodo de acercarse al nivel de activación cero. Cuanto mayor es el factor de decadencia, más fuerte es el mecanismo de decadencia. Siguiendo el razonamiento de Tsardias y Margaritis (1997, 1999), se propone como una función  $f$  de activación apropiada a la tangente hiperbólica (Haykin, 1994).

Bajo el paradigma de los *FCMs*, las influencias ejercidas sobre el nodo  $i$  por el resto de los nodos vecinos, se engloban en el término  $A_i$  en forma de *consistencias contextuales*, que se definen posteriormente, asumiendo que dichas consistencias podrían no ser simétricas. La auto-influencia ejercida por cada nodo sobre sí mismo se refleja en el término de memoria  $D_i(t)$ . Una vez definido el proceso de actualización de los estados de las neuronas a través de la ecuación (4.19), el objetivo se centra ahora en calcular: *a)* los pesos causales o coeficientes de consistencia  $w_{ik}(t)$  y *b)* el factor de decadencia.

#### *a.1) Cálculo de los pesos causales*

Como se ha mencionado previamente, el peso causal  $w_{ik}(t)$ , en cada iteración, se define como un coeficiente contextual como sigue. Teniendo en cuenta la relación entre la posición de un píxel  $(x,y)$  y el nodo  $i$  de la red, la vecindad  $N_i^m$  contiene los  $m$  nodos en los alrededores de  $i$  obtenidos a partir de la imagen original izquierda que originó el mapa de disparidad y posteriormente la estructura de la red; por tanto, esos  $m$  nodos representan las  $m$  posiciones espaciales conectadas con el píxel  $(x,y)$ . Un valor típico de  $m$  utilizado en la literatura es 8, que define una región  $3 \times 3$ ; este es el valor escogido en nuestro trabajo. Hemos probado otros valores mayores que 8, comprobando la existencia de una sobre-influencia de la vecindad a través de la ecuación (4.19). Como puede observarse a partir de dicha ecuación, el índice  $k$  varía de 1 a  $q$ , lo que significa que el nodo  $i$  puede recibir influencias de todos los nodos de la red. Sin embargo, desde el punto de vista práctico y bajo los principios de la escuela de psicología de la Gestalt, anteriormente mencionados, en el caso de las imágenes sólo los píxeles vecinos deben ejercer la influencia, y no los alejados espacialmente. Esto es lo que justifica la definición del entorno de vecindad y posteriormente la asignación de valores nulos a los pesos de conexión tal y como se define en la ecuación (4.20). En relación con la vecindad, para los nodos en los bordes de la imagen, la vecindad sólo incluye píxeles pertenecientes a la imagen, es decir,  $m = 3$  en las cuatro esquinas y  $m = 5$  en el resto de bordes.

Finalmente, definimos el coeficiente de consistencia en la iteración  $t$  a través de las ecuaciones (4.20) y (4.21).

$$w_{ik}(t) = \begin{cases} 1 - |D_i(t) - D_k(t)| & k \in N_i^m, \quad i \neq k \\ 0 & k \notin N_i^m, \quad i = k \end{cases} \quad (4.20)$$

$$w_{ik}(t) = \begin{cases} -w_{ik}(t) ; & \text{sgn}(w_{ik}(t)D_k(t)) \neq \text{sgn}(D_k(t)) \\ w_{ik}(t) ; & \text{e.c.c} \end{cases} \quad (4.21)$$

donde  $\text{sgn}$  representa la función matemática *signo*.

#### a.2) Obtención del factor de decadencia

Definimos el factor de decadencia basándonos en la hipótesis siguiente: si un nodo  $i$  ha alcanzado un alto grado de estabilidad y se desestabiliza por medio de este factor, dicho nodo volverá rápidamente de nuevo a la estabilidad. De esta forma lo que se pretende es asegurarnos de que el nodo es realmente estable y no se trata de una estabilidad ficticia. Esta idea procede de los trabajos de investigación realizados por Tsardias y Margaritis (1997, 1999).

La implementación de la idea anterior se materializa en nuestro trabajo de investigación como sigue. Construimos un acumulador de celdas de tamaño  $q = M \times N$ , y por tanto con la misma dimensión que la red, donde cada celda  $i$  se asocia al nodo de idéntico nombre. La celda  $i$  contiene el número de veces,  $h_i$ , que el nodo  $i$  ha cambiado significativamente su nivel de activación o estado. Inicialmente todos los valores  $h_i$  se ponen a cero, y posteriormente se hace  $h_i = h_i + 1$  si se cumple la siguiente relación:  $|D_i(t+1) - D_i(t)| > \varepsilon$ . Por tanto, la estabilidad del nodo  $i$  se mide como la fracción de cambios acumulados por la celda  $i$  en relación con los cambios producidos en su vecindad  $k \in N_i^m$  y en función del número de iteraciones  $t$ . Finalmente, teniendo en cuenta lo anterior, el factor de decadencia se define como sigue:

$$d_i = \begin{cases} 0 & h_i = 0 \text{ y } \overline{h_k} = 0 \\ \frac{h_i}{(\overline{h_k} + h_i)t} & \text{e.c.c} \end{cases} \quad (4.22)$$

donde  $h_i$  se definió más arriba y  $\overline{h_k}$  es el promedio de valores acumulados por los nodos vecinos  $k \in N_i^m$  a lo largo de las  $t$  iteraciones. Como puede observarse a partir de la

ecuación (4.22), si  $h_i = 0$  y  $\overline{h_k} = 0$ , el factor de decadencia toma el valor nulo, esto significa que no se producen cambios en los estados de la red, es decir, se ha alcanzado un grado de estabilidad alto. Cuando la fracción de cambios  $h_i$  es pequeña, la estabilidad del nodo  $i$  también es alta y el factor de decadencia tiende a cero. Por el contrario, si  $h_i$  es grande y  $\overline{h_k}$  pequeña, significa que el nodo  $i$  es inestable y sus vecinos estables, en este caso el factor de decadencia es alto. Si los dos valores, tanto  $h_i$  como  $\overline{h_k}$  son altos al mismo tiempo, el área donde se sitúan los píxeles que están interviniendo en el proceso corresponde a una zona de alta variabilidad, en cuyo caso el factor de decadencia tiende a cero a medida que aumenta  $\overline{h_k}$ , es decir, la desestabilización del nodo  $i$  a través del factor de decadencia carece de sentido puesto que el nodo es intrínsecamente inestable. Incluso si la parte de la fracción, excluyendo el término  $t$ , resultase ser constante, el factor de decadencia también se le hace tender a cero a medida que aumenta  $t$ . Esto quiere decir que quizás al inicio del proceso iterativo puedan ocurrir algunos cambios y posteriormente, a medida que el proceso progresa, no se detecten más cambios, lo cual es un signo inequívoco de estabilidad en el nodo. El factor de decadencia, por tanto, resta una fracción al nuevo nivel de activación obtenido con el primer término de la ecuación (4.19). Dada la definición de esta ecuación, podría ocurrir que durante la actualización del grado de activación de un nodo en cuestión, el nivel de activación tomara valores menores o mayores que  $-1$  ó  $+1$ , en cuyo caso el nivel se fijaría directamente a  $-1$  ó  $+1$ , respectivamente.

#### b) Resumen del proceso FCM

El proceso *FCM* en su conjunto se sintetiza como sigue:

- *Inicialización:* cargar cada nodo  $i$  con  $D_i(t=0)$ , que representa la disparidad normalizada al rango  $[-1, +1]$ , obtenida tras aplicar las restricciones epipolar, semejanza y unicidad, según se describe en la sección 4.2.5; fijar  $\varepsilon$  a 0.05 y  $t_{max}$  a 10 iteraciones. Estos valores se han establecido mediante prueba y error con un conjunto de imágenes patrón, suficientes para el conjunto de experimentos llevados a cabo en las imágenes estereoscópicas analizadas. Definir  $nc$  como el número de nodos que cambian sus valores de estado o niveles de activación en cada iteración.



- *Proceso FCM:*

$$t = 0;$$

$$\text{mientras } t < t_{\max} \text{ o } nc \neq 0 \text{ hacer}$$

$$t = t + 1;$$

$$nc = 0;$$

$$\text{para cada nodo } i$$

$$\text{actualizar } D_i(t+1) \text{ según la ecuación (4.19)}$$

$$\text{si } |D_i(t+1) - D_i(t)| > \varepsilon \text{ entonces}$$

$$nc = nc + 1;$$

$$\text{fin del si;}$$

$$\text{fin del para;}$$

$$\text{fin del mientras;}$$
- *Salidas:* los estados  $D_i(t)$  para todos los nodos actualizados, a partir de los cuales se aplica la transformación inversa para devolver los valores al rango  $[0, \theta_{\max}]$ , obteniendo el mapa de disparidad final refinado.

#### 4.2.6.2 Red Neuronal de Hopfield (HNN)

##### a) Revisión de la Red Neuronal de Hopfield

El paradigma *HNN* propuesto inicialmente por Hopfield y Tank (1985, 1986) ha sido ampliamente utilizado para resolver problemas de optimización. En tales problemas se requiere fijar dos características inherentes a la propia red (Joya y col., 2002): *a)* su dinámica de activación y *b)* una función de energía asociada, que disminuye a medida que la red evoluciona. En la figura 4.8 se representa un modelo esquemático de este tipo de redes donde los círculos numerados representan los nodos de la red. Los nodos se encuentran conectados entre sí por enlaces, donde cada uno de estos enlaces tiene asociado un peso de conexión  $T_{ij}$ , que expresa de algún modo un tipo de conexión existente entre un nodo  $i$  y otro  $j$ . Por otra parte, cada nodo  $i$  recibe del exterior una entrada  $U_i$ .

La *HNN* es una red recurrente donde cada nodo se encuentra conectado a todos los demás, pero no consigo mismo, es decir, no existen conexiones auto-recurrentes. La red representa un modelo dinámico, donde las salidas de los nodos se utilizan como realimentación para el resto de nodos, induciendo nuevas respuestas de los nodos realimentados. Por tanto, se trata de un modelo dinámico que evoluciona con el número de iteraciones, repitiéndose el proceso una y otra vez hasta conseguir que los cambios en

las salidas entre iteraciones consecutivas sean cada vez más y más pequeños, hasta conseguir que las salidas de los nodos no varíen entre iteraciones sucesivas, en cuyo caso se dice que se ha alcanzado la estabilidad de la red.

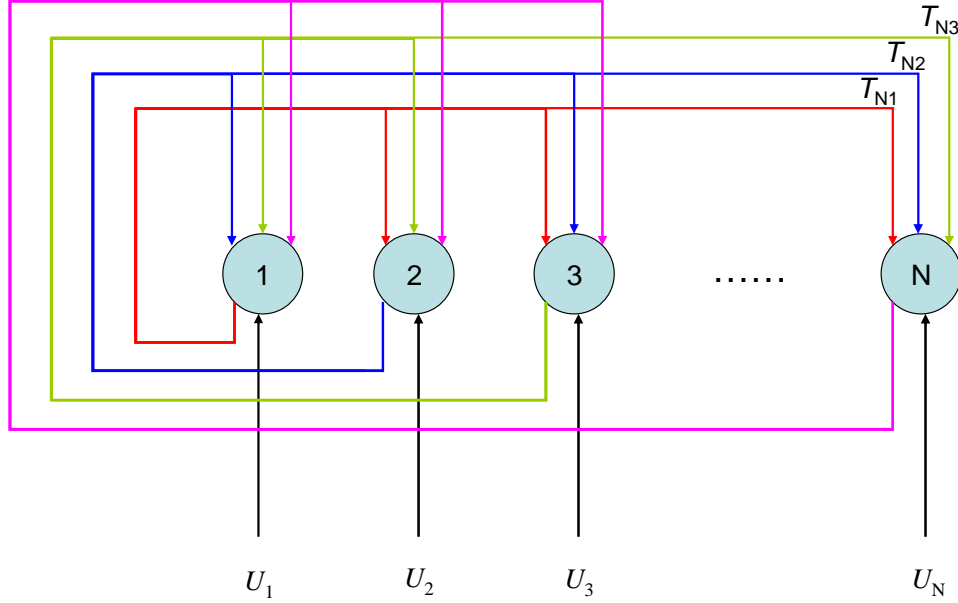


Figura 4.8 Modelo de la Red Neuronal de Hopfield

Los pesos de conexión entre los nodos de la red pueden organizarse en forma de una matriz  $T$ . Cohen y Grossberg (1983) y Qiao y col. (2001) demuestran que una red recurrente de esta naturaleza es estable si la matriz  $T$  es simétrica con ceros en su diagonal principal, esto es, cuando  $T_{ik} = T_{ki}$  para toda  $i$  y  $k$  y  $T_{ii} = 0$  para todos los nodos en la red. Aunque algunos estudios llevados a cabo por Lee y Chuang (2005) en este tipo de redes neuronales han determinado que el problema de optimización puede resolverse mediante memorias asociativas asimétricas óptimas, en este trabajo hemos optado por elegir el enfoque clásico. Por tanto, la dinámica de la red se establece considerando la matriz  $T$  simétrica.

Como se ha puesto de manifiesto previamente, la entrada al nodo  $i$ -ésimo proviene de dos fuentes: aportaciones externas y aportaciones internas provenientes de otros nodos. La entrada total  $u_i$  al nodo  $i$  resulta ser la siguiente:

$$u_i(t) = \sum_{k \neq i} T_{ik} D_k(t) + U_i(t) \quad (4.23)$$

donde el valor  $D_k(t)$  representa la salida del nodo  $k$ -ésimo en la iteración  $t$ ;  $T_{ik}$  es el peso

de la conexión entre nodos  $i$  y  $k$ , perteneciente a la matriz  $T$ , y  $U_i$  representa un valor prefijado de entrada externo que se utiliza para establecer el nivel general de excitabilidad de la red, según el esquema de la figura 4.8.

Existen dos tipos de redes de Hopfield (Joya y col., 2002; Haykin, 1994), a saber:

- 1) modelos analógicos en los que a los estados de las neuronas se les permite variar de forma continua en un intervalo, tal como el  $[-1, +1]$ ;
- 2) modelos discretos en los que estos estados se limitan a los valores binarios  $-1$  y  $+1$ .

El inconveniente del modelo binario estriba en el hecho de que los valores de un nodo pueden oscilar entre los dos estados binarios alternativamente, produciéndose cambios bruscos en el estado general de la red. Hopfield ha demostrado que las redes analógicas funcionan mejor que las binarias ya que tienen la capacidad de suavizar la superficie de la función de energía, evitando así que el sistema quede atascado en mínimos locales (Hopfield y Tank, 1985, 1986). Por este motivo, en el presente trabajo de investigación, se ha optado por el modelo analógico.

Para redes analógicas la entrada total a un nodo se convierte en un valor de salida por una función de activación monótona, preferiblemente de tipo sigmoideal, en lugar de la operación umbral que se utiliza para redes discretas de Hopfield (Qiao y col., 2001). La dinámica de un nodo dado  $i$  se define mediante la ecuación diferencial siguiente:

$$\frac{du_i}{dt} = -\frac{u_i}{R_i} + \sum_{k \neq i} T_{ik} D_k(t) + U_i \quad \text{donde} \quad D_k(t) = g(u_k) \quad \forall k \quad (4.24)$$

donde  $g(u_i)$  es la función de activación sigmoideal, anteriormente mencionada, y  $R_i$  es una constante de tiempo que puede fijarse a 1 por simplicidad (Haykin, 1994; Yu y Tsai, 1992). En el modelo presentado en este trabajo hemos escogido como función de activación sigmoideal la función tangente hiperbólica (Haykin, 1994),  $g(u_k) = \tanh(u_k / \beta)$ . Las razones para su elección se basan en el hecho de que esta función es diferenciable, suave y monótona, lo que contribuye a la estabilidad de la red

(Qiao y col. 2001). Una discusión detallada sobre la configuración del intervalo de tiempo  $dt$  y la ganancia  $\beta^{-1}$  puede encontrarse en Joya y col. (2002). A medida que  $dt$  aumenta, la probabilidad de que la energía caiga en un mínimo local también aumenta. Según algunos experimentos llevados a cabo por Joya y col. (2002) donde este parámetro ha sido fijado a valores en el rango 1 a  $10^{-2}$ , el mejor resultado se alcanza con el mínimo valor, es decir con  $10^{-2}$ , por tanto, bajo la suposición de que valores mínimos pueden favorecer el proceso y evitar la caída en mínimos locales, en este trabajo hemos fijado dicho parámetro a  $10^{-3}$  que es un orden de magnitud menor que el experimentado en Joya y col. (2002). La manera de evitar que una red continua caiga en mínimos locales y hacer que la red converja a un estado solución es disminuir  $\beta$  a lo largo del proceso de optimización, hasta llegar en sucesivas iteraciones a que  $\beta$  alcance el valor cero. Esta estrategia recuerda claramente los procesos de enfriamiento simulado donde se comienza por una temperatura elevada, que va disminuyendo (enfriamiento) a medida que el sistema progresa. De esta manera, comenzando con un  $\beta$  suficientemente alto, la red evoluciona hasta alcanzar un estado estable (que no necesariamente es una solución), después  $\beta$  disminuye y la red progresa de nuevo hacia un nuevo estado más estable, y así sucesivamente. El proceso finaliza cuando  $\beta$  llega a cero, en ese momento el estado estable alcanzado debería corresponderse con un mínimo global de energía.

De acuerdo con los resultados presentados en Kasetkasem y Varshney (2002) y Starink y Backer (1995), hemos propuesto para esta investigación la estrategia de planificación siguiente  $\beta(t) = \beta_0 / \log(t+1)$  donde  $t$  representa el número de iteración. El valor inicial  $\beta_0$  se calcula como sigue (Laarhoven y Aarts, 1989):

- 1) Seleccionamos cuatro pares de imágenes estereoscópicas, donde los nodos han sido inicializados previamente a los valores  $D_i$  dados por el proceso de correspondencia. En este estado se calcula la energía inicial de la red.
- 2) Escogemos un  $\beta$  inicial, de forma tal que permita que alrededor del 80% de todas las transiciones sean aceptadas, es decir, transiciones que disminuyen la función de energía. Este valor inicial se modifica hasta conseguir dicho porcentaje.
- 3) Calculamos las  $M$  transiciones  $\Delta E_i$  así obtenidas y buscamos un valor de  $\beta$  para

el cual  $\frac{1}{M} \sum_{k=1}^M \exp(-\Delta E_k / \beta) = 0.8$ , después de rechazar los términos de orden superior de la expansión de Taylor de la exponencial anterior,  $\beta = 5 \langle \Delta E_k \rangle$ , donde  $\langle \cdot \rangle$  representa el valor medio. En nuestros experimentos hemos obtenido  $\langle \Delta E_k \rangle = 0.87$ , consiguiendo así un valor  $\beta_0 = 4.35$ . En el trabajo de Starink y Backer (1995) se utiliza una planificación de enfriamiento simulado con  $\beta_0 = 2$ , es decir, del mismo orden de magnitud que en nuestro caso. Teniendo en cuenta que  $\beta(t) = 0$ ,  $t \rightarrow +\infty$  y considerando  $t = 10^{10}$ , obtenemos  $\beta = 0.43$ , es decir,  $\beta^{-1} = 2.30$ . En nuestro enfoque de correspondencia estereoscópica hemos llevado a cabo diferentes experimentos mediante el proceso de planificación descrito anteriormente. Además hemos supuesto que la ganancia se mantiene constante a lo largo de todo el proceso, verificando que en ambos casos, los resultados finalmente obtenidos son totalmente equivalentes, sin que el proceso de enfriamiento llegue a conseguir una mejora aceptable. Por consiguiente hemos optado por fijar la ganancia al valor de 2.30 durante el proceso completo.

El modelo proporcionado en la ecuación (4.24) se corresponde con el circuito clásico de Hopfield (Hopfield y Tank, 1985, 1986; Kosko, 1992) que proviene de los sistemas dinámicos de Cohen-Grossberg (Cohen y Grossberg, 1983). En Kosko (1992) se ha demostrado la estabilidad global de estos sistemas bajo el supuesto de positividad  $dg/dt > 0$  y considerando que el coeficiente que aparece en el término izquierdo de la ecuación (4.24), toma siempre el valor unidad y por tanto positivo, esto es, si dicho término se reescribe como  $C \frac{du_i}{dt}$ , es obvio que  $C$  siempre toma el valor unidad. Puesto que  $g$  se ha definido como la función tangente hiperbólica, la primera condición también es verdadera. Además, a la estabilidad global también contribuye el hecho de que  $U_i$  varía lentamente, por tanto, puesto que este valor se obtiene según la ecuación (4.31) a partir del nivel de activación del nodo en cuestión, y dado que el diseño elegido para el modelo de red neuronal de Hopfield no permite una variación brusca en los niveles de activación, tal y como se concluye en la explicación relativa a su cómputo a través de dicha ecuación, esto nos conduce a la conclusión de que, en efecto, la variación de  $U_i$  es relativamente lenta. La estabilidad de la red neuronal también ha sido estudiada bajo

diferentes perspectivas en Qiao y col. (2001) o Zhao (2004). Por tanto, pertenece a la importante clase de modelos de redes neuronales realimentadas que son globalmente estables.

La magnitud que describe el estado de la red recibe el nombre de *energía*, y se define de la siguiente manera:

$$E(t) = -\frac{1}{2} \sum_i \sum_{k \neq i} T_{ik} D_i(t) D_k(t) - \sum_i U_i D_i(t) + \beta \sum_i \int_0^{D_i} g^{-1}(D) dD \quad (4.25)$$

Según los resultados presentados en Joya y col. (2002), el término integral en (4.25) está limitado por  $\beta \ln 2 \approx 0.19$  cuando  $D_i(t)$  es  $+1$  o  $-1$  y toma el valor nulo cuando  $D_i(t)$  es cero. En nuestros experimentos hemos verificado que este término no contribuye a la estabilidad de la red, y la energía solamente varía en cantidades muy pequeñas en relación a los otros dos términos en (4.25), llegando a alcanzar en la mayoría de los casos cantidades prácticamente despreciables. Por tanto, y por simplicidad, desde el punto de vista práctico hemos eliminado el término integral de la ecuación (4.25).

En definitiva, el modelo de Hopfield continuo descrito por el sistema de la ecuación diferencial de primer orden no lineal (4.24) representa una trayectoria en el espacio de fase, que busca los mínimos de la función de energía en (4.25). A continuación centramos el problema en la obtención de la función de energía para nuestro problema basado en la aplicación de la continuidad de la disparidad. Este proceso pasa por combinar los valores de los estados de los nodos entre un nodo central  $i$  y sus vecinos  $k$ , es lo que denominamos, por comparación con el caso de los *FCMs*, influencias contextuales o simplemente consistencias, obtenidas desde dos puntos de vista que denominamos respectivamente *consistencia* e *información contextual mutua*. Además de estas relaciones mutuas es necesario considerar la propia información contenida en el propio nodo (píxel)  $i$ , que denominamos *autoinformación de los datos* siguiendo la nomenclatura de la literatura especializada. A través de la consistencia e información contextual mutua, junto con la autoinformación, se obtienen tanto los pesos de conexión entre los nodos como las entradas externas de la red, completando así la definición del modelo de red de Hopfield.

## b) Cómputo de consistencias e información contextual mutuas

### b.1) *Consistencia de los datos*

Durante el proceso de optimización los estados iniciales  $D_i(t)$  se modifican intentando alcanzar la estabilidad de la red. Ahora el objetivo es obtener la consistencia de los datos entre los nodos  $i$  y  $k$  mediante la definición de lo que denominamos *coeficiente de consistencia*  $w_{ik}(t)$ , en cada iteración  $t$ . Dado el nodo  $i$ , se considera su vecindad  $m$ -conectada  $N_i^m$ , como viene siendo habitual bajo el criterio de agrupación establecidos por los principios de proximidad y conectividad de acuerdo con la filosofía de la Gestalt (Wang, 2005), donde  $m$  podría valer 4, 8, 24, 48 o cualquier otro valor teniendo en cuenta solamente direcciones horizontales, verticales o diagonales. Si bien, como en el caso de los *FCMs* y por las mismas razones expresadas allí, se ha elegido el valor típico de 8, que corresponde a un píxel central y sus 8 vecinos.

Para cada nodo  $i$ , sólo se pueden establecer consistencias entre nodos  $k$ , donde  $k \in N_i^m$  e  $i \neq k$ , en caso contrario si  $k \notin N_i^m$  se asume que no se define consistencia entre los nodos  $i$  y  $k$ . Esto se justifica bajo la hipótesis de que sólo se pueden establecer relaciones locales entre píxeles con disparidades similares. Dados dos nodos  $i$  y  $k$ , se dice que  $k \in N_i^m$  es consistente con  $i$  si ambos simultáneamente contienen información de los datos similares, es decir, valores de activación similares y por tanto disparidades también similares. En caso contrario se dice que son inconsistentes. La consistencia de datos entre los nodos  $i$ ,  $k$ , se obtiene mediante el denominado coeficiente de consistencia como sigue:

$$w_{ik}(t) = \begin{cases} 1 - |D_i(t) - D_k(t)| & k \in N_i^m, \quad i \neq k \\ 0 & k \notin N_i^m, \quad i = k \end{cases} \quad (4.26)$$

A partir de (4.26) se puede observar que  $w_{ik}(t)$  varía en el rango  $[-1, +1]$ . La influencia ejercida por el nodo  $k$  sobre el nodo  $i$  será positiva (refuerzo) o negativa (penalización). Una consistencia positiva de los datos contribuirá a la estabilidad de la red.

### b.2) Consistencia procedente de la información contextual

En algunos trabajos existentes en la literatura donde las imágenes constituyen el eje central de los mismos, tal como en Kasetkasem y Varshney (2002), se describe la interdependencia de píxel mediante la definición de un tipo de consistencia que se alcanza bajo la consideración de información contextual. Haciendo uso de este concepto lo trasladamos a nuestra propuesta de red *HNN*. Así, dado el nodo  $i$  en la posición del píxel  $(x,y)$  con un nivel de activación o estado  $D_i$  y un conjunto de nodos  $k \in N_i^m$  con valores de estado  $D_k$ , una medida de la consistencia contextual entre el nodo  $i$  y sus  $k$  vecinos se puede expresar como:

$$E_i(t) = \sum_{k \in N_i^m} D_i(t) D_k(t) \quad (4.27)$$

Es evidente que este término representa una interrelación entre el estado de los nodos de la red. También representa un tipo de influencia externa ejercida por los  $k$  nodos sobre el nodo  $i$ . Como  $D_i(t)$  y  $D_k(t)$  varían en el rango  $[-1,+1]$ , dado  $D_i(t)$ , el término  $E_i(t)$  será máximo cuando los  $D_k(t)$  valores estén próximos a  $D_i(t)$ . Es más, asumiendo la relación de vecindad ocho, cuando  $D_i(t)$  y  $D_k(t)$  toman simultáneamente valores de  $+1$  o  $-1$ ,  $E_i(t)=8$ , es decir, alcanza su máximo valor. Por el contrario, si  $D_i(t)=+1$  y todos los vecinos toman  $D_k(t)=-1$  o a la inversa  $D_i(t)=-1$  y todos sus vecinos  $D_k(t)=+1$ , entonces en ambos casos  $E_i(t)=-8$ , es decir, este término adquiere su mínimo valor. Cabe señalar que (4.27) puede considerarse como una implementación del potencial de Gibbs en una vecindad dada bajo el criterio de los denominados *MRFs* que se estudiaron en la sección 2.2.3.1 (Kasetkasem y Varshney, 2002).

Una vez que los datos y las consistencias contextuales se han especificado, buscamos una función de energía tal que tome un valor bajo cuando ambas consistencias sean altas y viceversa, que el valor sea alto cuando las consistencias sean bajas. Este término de energía se expresa como:

$$E_c(t) = -\frac{A}{2} \sum_i \sum_{k \in N_i^m} \{ [\text{sgn}(w_{ik}(t))]^{v+1} w_{ik}(t) - \delta_{ik} \} D_i(t) D_k(t) \quad (4.28)$$

$$\text{donde} \quad \text{sgn}(w_{ik}(t)) = \begin{cases} -1 & \text{si } w_{ik}(t) \leq 0 \\ +1 & \text{si } w_{ik}(t) > 0 \end{cases} \quad y \quad \delta_{ik} = \begin{cases} 1 & \text{si } i = k \\ 0 & \text{si } i \neq k \end{cases}$$



donde  $A$  es una constante positiva que se define después,  $sgn$  representa como siempre la función matemática *signo*; finalmente  $v$  representa el número de valores negativos en el conjunto  $C \equiv \{w_{ik}(t), D_i(t), D_k(t)\}$ . Así, dado  $S \equiv \{s \in C / s < 0\} \subseteq C$ ,  $v = \text{card}(S)$  siendo  $\text{card}$  la función matemática *cardinal*;  $\delta_{ik}$  se introduce en este término para anular la propia contribución del nodo  $i$  en el cómputo del valor de la energía. Esto se hace así porque dicha contribución se realiza específicamente a través del término obtenido mediante la autoinformación de los datos, evitando de este modo una excesiva autocontribución.

La tabla 4.1 muestra un ejemplo pedagógico sobre el comportamiento del término energético  $E_C(t)$  frente a las consistencias de los datos y contextuales. Como puede fácilmente deducirse, la energía disminuye en la misma medida en que los datos originales (disparidades) que representan en realidad los estados de las neuronas, son ambos simultáneamente consistentes (filas 1 y 4 en la parte izquierda de la tabla 4.1); en caso contrario bajo la ausencia de consistencia, el término energético correspondiente aumenta. Hemos creído conveniente considerar el hecho de que las consistencias de los datos tienen mayor prioridad que las contextuales, puesto que han sufrido menos manipulación durante el proceso; de modo que bajo este criterio si  $w_{ik}(t) < 0$  entonces la energía aumenta.

Tabla 4.1 Comportamiento del término energía frente a los datos y las consistencias contextuales

$w_{ik}(t)$	$D_i(t)$	$D_k(t)$	$E_C(t)$	$w_{ik}(t)$	$D_i(t)$	$D_k(t)$	$E_C(t)$
+	+	+	−	−	+	+	+
+	+	−	+	−	+	−	+
+	−	+	+	−	−	+	+
+	−	−	−	−	−	−	+

### b.3) Autoinformación de los datos

Hemos analizado las interrelaciones entre los nodos en una vecindad determinada, en base a los datos y las consistencias contextuales. Esto implica que el estado o nivel de activación de cada nodo evoluciona de acuerdo con la información proporcionada por los nodos vecinos, ignorando hasta ahora su propia información. Esto puede conducir a un estado incorrecto para el nodo en cuestión. Para superar este inconveniente suponemos que cada nodo debe contribuir a la evolución de su propio

estado a través de la información que él mismo pueda proporcionar. La propia información de los datos se modela como una especie de auto-consistencia basada en la hipótesis de que un nodo en la red con un valor de estado (disparidad)  $D_i(t)$  debe actualizarse teniendo en cuenta este valor, ya que el mismo se ha obtenido a partir de los datos originales. Para conseguir esto se define el siguiente término energético:

$$E_B(t) = -B \sum_i D_i(t) D_i(t) \quad (4.29)$$

La constante  $B$  en la ecuación (4.29) es un número positivo que se define más adelante. De este modo, independientemente de si  $D_i$  es positiva o negativa, el producto  $D_i(t)D_i(t)$  es siempre positivo, y el término  $E_B(t)$  en cada iteración es mínimo, como cabe esperar bajo la consideración de que la energía de la red debe tender hacia valores mínimos para llegar a configuraciones de estados estables.

#### *b.4) Derivación de los pesos de conexión y las entradas externas para la HNN*

Asumiendo las consistencias contextuales y de los datos, ecuación (4.28), y la información propia de los datos, ecuación (4.29), obtenemos la función de energía dada en la ecuación (4.30), que engloba los términos energéticos previos. Esta es la energía a minimizar mediante el proceso de optimización bajo el paradigma establecido en el modelo *HNN*.

$$E(t) = E_C(t) + E_B(t) = -\frac{A}{2} \sum_i \sum_{k \in N_i^m} \left\{ \left[ \text{sgn}(w_{ik}(t)) \right]^{v+1} w_{ik}(t) - \delta_{ik} \right\} D_i(t) D_k(t) - B \sum_i D_i(t) D_i(t) \quad (4.30)$$

Comparando las expresiones (4.25) y (4.30) sin el término integral en (4.25), resulta fácil identificar los pesos de conexión y las entradas externas como sigue:

$$T_{ik} = A \left[ \text{sgn}(w_{ik}(t)) \right]^{v+1} w_{ik}(t) - \delta_{ik}; \quad U_i(t) = B D_i(t) \quad (4.31)$$

De acuerdo con la discusión en el punto *a)* anterior de esta misma sección, para garantizar la convergencia hacia un estado estable (Cohen y Grossberg, 1983), se requieren pesos de interconexión simétricos y con auto-realimentación nula, por tanto, resulta fácil deducir que mediante la asignación  $A = B = 1$  ambas condiciones pueden fácilmente deducirse de (4.30). Además, anteriormente establecíamos que una de las

condiciones de estabilidad de la red provenía del hecho de asegurar que la entrada externa  $U_i(t)$  debe variar lentamente. Pues bien, esto en cierta medida se consigue gracias a que la red se carga inicialmente con el mapa de disparidad proporcionado por cualquiera de los métodos de decisión elegidos, bien sea el *CFI*, *SFI*, *DES* o *MCDMF* y por tanto, la red comienza con un cierto grado de estabilidad haciendo que los valores de activación de los nodos cambien lentamente durante el proceso de optimización. Además, gracias a la elección del entorno de vecindad con su limitación espacial, sólo un pequeño número de nodos están interconectados, en nuestro caso nueve, lo que contribuye ciertamente a la estabilidad global de la red como se demuestra en los estudios realizados en Zhao (2004).

La ecuación de la energía en (4.30) representa un compromiso entre los datos y la información contextual proveniente del entorno espacial que rodea el nodo  $i$ , y a la vez la propia información suministrada por el nodo  $i$ . Las constantes  $A$  y  $B$  pueden fijarse de modo que definan una mayor o menor influencia de cada uno de los términos involucrados en (4.30). Sin embargo, diversos experimentos llevados a cabo en el conjunto de imágenes disponibles, nos conduce a que con la elección realizada igualando las influencias entre los términos por medio de los valores  $A$  y  $B$  iguales a la unidad, es apropiada, no observándose mejoras significativas salvo que cuando cualquiera de ellos tiende a anularse los resultados empeoran considerablemente.

La ecuación (4.24) describe la evolución en el tiempo de la red a través de los estados de sus nodos. Los estados de los nodos, tanto del central como de los vecinos  $D_i(t)$ ,  $D_k(t)$  se calculan resolviendo la ecuación diferencial (4.24) mediante el método Runge-Kutta. Como fácilmente puede deducirse, la energía en (4.30) se obtiene considerando los valores de estado y una especie de atracción mutua derivada de ambos, datos y consistencias contextuales. La obtención de una función de energía involucrando atracciones entre puntos fijos ha sido un aspecto bien estudiado en el trabajo de Muezzinoglu y col. (2005) para memorias asociativas de Hopfield, preservando pesos simétricos y sin auto-realimentación. Por tanto, podemos suponer que bajo la atracción de los datos y las consistencias contextuales, consideradas en este trabajo de investigación, nuestro enfoque basado en el modelo analógico de Hopfield actúa adecuadamente.

c) Resumen del proceso HNN

Teniendo en cuenta la función de energía definida para describir el comportamiento de la red neuronal de Hopfield, el filtrado del mapa de disparidad inicial se logra dejando que la red evolucione hasta conseguir un estado estable, es decir, cuando no ocurra ningún cambio en los estados de sus nodos durante el procedimiento de actualización. El proceso de suavizado completo puede resumirse como sigue:

- *Inicialización:* se crea un nodo  $i$  para cada posición de píxel  $(x,y)$  de la imagen izquierda;  $t = 0$  (número de iteración); se carga cada nodo con el nivel de actividad o valor de estado  $D_i(t)$  definido en la sección 4.2.6.2a; se calculan  $T_{ij}$  y  $U_i(t)$  mediante la ecuación (4.31); se fija  $\varepsilon$  al valor 0.05, como constante para acelerar la convergencia y  $t_{max}$  a 10, que representa el máximo número de iteraciones permitido; se establecen los valores de las constantes como sigue:  $R_i = 1$ ;  $\beta = 0.43$ ;  $dt = 10^{-3}$ . Finalmente, se define  $nc$  como el número de nodos que cambian su valor de estado en cada iteración.
- *Proceso HNN:*
  - mientras*  $nc \neq 0$  *y*  $t < t_{max}$  *hacer*
    - $t = t + 1$ ;
    - $nc = 0$ ;
    - para* cada nodo  $i$ 
      - calcular  $u_i(t)$  empleando el método Runge-Kutta según (4.24)
      - actualizar  $D_i(t)$  según (4.24)
      - si*  $|D_i(t) - D_i(t-1)| > \varepsilon$  *entonces*
        - $nc = nc + 1$ ;
      - fin del si*;
    - fin del para*;
    - fin del mientras*;
- *Salidas:* los estados  $D_i(t)$  para todos los nodos actualizados, a partir de los cuales se aplica la transformación inversa para devolver los valores al rango  $[0, \theta_{max}]$ , obteniendo el mapa de disparidad refinado.

#### 4.2.7 Resumen de parámetros en el proceso basado en píxeles

En la tabla 4.2 se recogen los valores de los parámetros involucrados en el proceso de correspondencia basado en píxeles.

Tabla 4.2 Parámetros y umbrales implicados en el proceso de correspondencia basado en píxeles

Fase	Parámetro/Umbrales	Descripción
Restricción de unicidad	$U = 0.7$	Umbral para garantizar que las correspondencias sean suficientemente fiables. Paso 5 de los métodos <i>CFI</i> , <i>SFI</i> , <i>DES</i> y <i>MCDMF</i> , sección 4.2.4.
Restricción continuidad de la disparidad	$\varepsilon = 0.05$ $t_{max} = 10$	Parámetros involucrados en la inicialización de los procesos <i>FCM</i> y <i>HNN</i> , secciones 4.2.6.1 y 4.2.6.2, respect. $\varepsilon$ es una constante para acelerar la convergencia, y $t_{max}$ el máximo número de iteraciones permitido.
	$R_i = 1$ $\beta = 0.43$ $dt = 10^{-3}$	Parámetros involucrados en la inicialización del proceso <i>HNN</i> , sección 4.2.6.2. $R_i$ es una constante de tiempo, $\beta$ es la ganancia y $dt$ el intervalo de tiempo.

## 4.3 Correspondencia basada en regiones

### 4.3.1 Diseño del proceso

En la sección previa hemos considerado el procedimiento de diseño relativo a la correspondencia a nivel de píxel. Ahora ha llegado el momento de abordar el problema de la correspondencia basada en regiones, completando así los dos enfoques que se proponen en este trabajo de investigación, tal y como se ha puesto de manifiesto a lo largo de los capítulos precedentes así como en la introducción de éste. El diseño de este proceso está basado íntegramente en el trabajo de Herrera y col. (2009d).

Continuando con el esquema de la figura 3.2, y teniendo en cuenta que en este momento el proceso de segmentación ha finalizado, tenemos disponibles sendos conjuntos de regiones procedentes de cada una de las imágenes del par estereoscópico. Cada región identifica uno de los troncos existentes en las imágenes originales, que se describe por un conjunto de atributos tal y como se detalló en la sección 3.3.3 del capítulo tres. Estos atributos son los que reproducimos aquí por simplicidad: área de la región en píxeles, localización espacial del centroide en coordenadas cartesianas y polares, incluyendo el ángulo del mismo en estas últimas coordenadas. El ángulo constituye un parámetro fundamental en este proceso de correspondencia, como se verá posteriormente, dada la naturaleza geométrica circular de las imágenes. Finalmente, para cada región se obtienen los siete momentos invariantes de Hu.

Siguiendo el mismo esquema mantenido hasta el momento para el caso de los píxeles, la aplicación de las restricciones estereoscópicas, sección 4.1, conjuntamente con los atributos de las regiones, definidos previamente, permiten llevar a cabo el proceso. De forma más precisa, las restricciones en cuestión aplicadas en este enfoque son: *epipolar*, *semejanza*, *orden posicional* y *unicidad*. En lo que sigue, describimos la relación entre restricciones estereoscópicas y atributos de regiones.

### 4.3.2 Epipolar: centroide

El centroide de cada región, considerado como atributo de la misma, se utiliza como elemento clave para su aplicación bajo la restricción epipolar, sirviendo como guía para la correspondencia, tal y como se explica seguidamente.

La geometría del sensor define las líneas epipolares. El esquema de dicha geometría es el descrito en la sección 4.2.2, por tanto, las ecuaciones (4.1) a (4.4) que lo definen, se aplican de igual manera en este caso. Tal y como se define allí se pueden obtener distancias a puntos específicos. Así, es posible determinar la distancia del dispositivo de medida estereoscópica al centroide de las regiones, de esta forma se puede obtener la distancia al árbol que representa. Con esta distancia ya se pueden tomar decisiones tales como la de su inclusión en la parcela de muestra bajo estudio y a partir de aquí se puede obtener la densidad de árboles en dicha parcela. Calculando el ángulo central cubierto por la anchura del tronco y junto con la distancia desde el dispositivo, se pueden obtener medidas tales como el grosor de los troncos a diferentes alturas del suelo. Otra de las posibilidades que se ofrecen es la de calcular distancias desde el dispositivo a puntos en la base y la copa de un determinado árbol. Con estas distancias y utilizando los ángulos de proyección  $\alpha_i$  obtenidos a través de la ecuación (4.1) para dichos puntos, podemos calcular la altura del árbol aplicando reglas trigonométricas tales como el teorema del coseno. Siguiendo con este razonamiento, es posible calcular distancias a otros puntos significativos y a partir de ellos medir otras variables como las descritas en el capítulo uno.

Como ya se mencionó anteriormente, en las imágenes capturadas con lentes de ojo de pez y por la geometría del sistema, las líneas epipolares no son circunferencias concéntricas. Este hecho debe ser naturalmente considerado para la correspondencia estereoscópica de características, en este caso basado en regiones. La figura 4.9(b)

muestra seis líneas epipolares en la imagen derecha, que han sido generadas por los seis píxeles localizados en las posiciones marcadas con los cuadrados, que representan sus posiciones equivalentes en la imagen izquierda, figura 4.9(a). Dado un cuadrado, marcado en rojo en la imagen izquierda, representando el píxel en su misma localización, siguiendo la línea epipolar definida en la imagen derecha y en dirección Sur, debemos encontrar su píxel homólogo, figura 4.9(b). Aplicando este razonamiento para el píxel que representa al centroide de una región en la imagen izquierda, su correspondiente centroide en la imagen derecha debe situarse con toda probabilidad en la línea epipolar. Debido a que el propio dispositivo de medida puede introducir errores por una calibración incorrecta, hemos considerado un desplazamiento fuera de las líneas epipolares cuantificado en 10 píxeles de distancia, que permite ampliar el área de búsqueda con base en la línea epipolar. Además, sobre la línea epipolar o en su ampliación, los centroides homólogos se encuentran separados un cierto ángulo, medido en diferencias de ángulo en coordenadas polares, tal y como podemos observar en la figura 4.9(b) mediante la representación en cuadrados rojos y azules. Tras diferentes experimentos realizados con el conjunto de imágenes de prueba, la máxima separación en grados encontrada para este tipo de imágenes ha sido cuantificada en 22°.

Obviamente, el razonamiento previo se ha realizado considerando la imagen izquierda como la de referencia; si cambiamos la referencia a la derecha, la búsqueda de las correspondencias ahora sobre la izquierda se realiza en dirección opuesta, hacia el Norte, naturalmente siguiendo las mismas líneas epipolares.

Tal y como se puso de manifiesto en el caso del método de correspondencia basado en píxeles, en el trabajo de Elias (2007) se realiza una propuesta interesante relativa al proceso de correspondencia estereoscópica. En efecto, en él se formula la hipótesis de que cuando dos características, una de la imagen izquierda y otra de la derecha, tratan de emparejarse, lo lógico es que exista una preferencia mutua clara, esto es, la característica de la imagen izquierda prefiere a la de la derecha y viceversa. Como fácilmente puede deducirse, se trata de un hecho extraído de la vida social misma. Pues bien, esto nos permite aplicar esta idea en nuestra propuesta de correspondencia en el caso de las regiones. Ahora los puntos de referencia lo constituyen los centroides de las regiones bajo correspondencia, de suerte que dado el centroide de una región en la imagen izquierda buscamos su centroide homólogo en la imagen derecha siguiendo las

líneas epipolares; del mismo modo también a la inversa, esto es, dado un centroide en la imagen derecha, buscamos en sentido contrario en la imagen izquierda, siguiendo también las líneas epipolares. En aplicación del principio de preferencias mutuas planteados en Elias (2007) o en Ogale y Aloimonos (2005), las regiones representadas por esos centroides deben tener preferencias mutuas, tal y como se determina posteriormente.

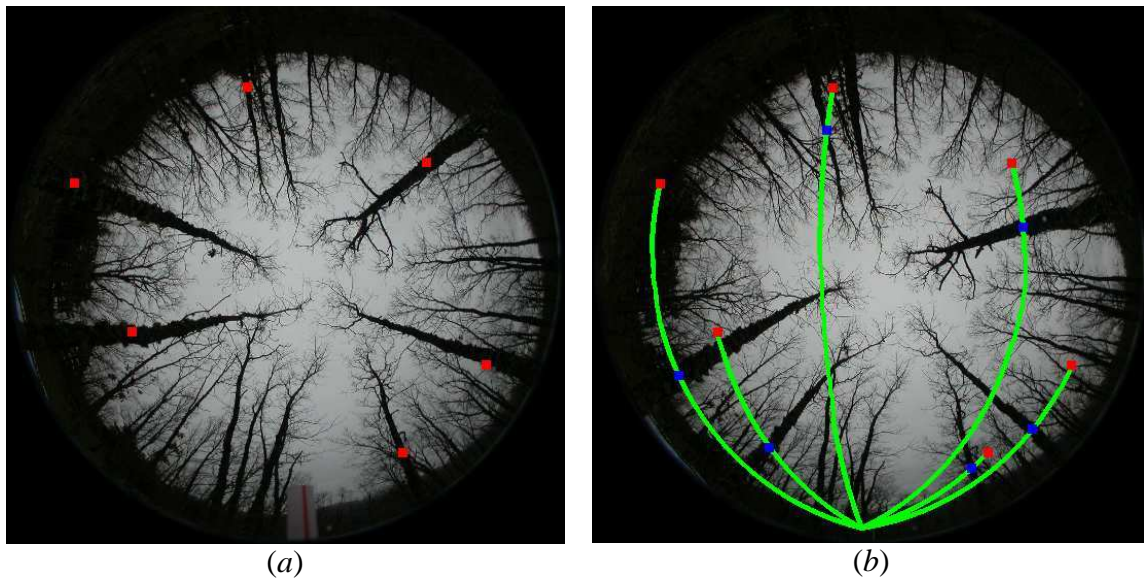


Figura 4.9. (a) Imagen izquierda señalada mostrando seis cuadrados representativos de otros tantos píxeles. (b) Líneas epipolares en la imagen derecha generadas a partir de las localizaciones en la imagen izquierda señaladas por los cuadrados.

### 4.3.3 Semejanza: áreas y momentos invariantes de Hu

Como se mencionó anteriormente, cada región en las imágenes izquierda y derecha del par estéreo tiene su propio conjunto de propiedades. Los momentos de Hu presentan la particularidad, bien conocida en el mundo de la Visión por Computador, de que son invariantes a traslaciones, rotaciones y cambios de escala, tal y como se puso de manifiesto en la sección 3.3.3.

De particular interés resulta la invarianza a rotaciones en el caso de las imágenes bajo procesamiento, ya que los troncos aparecen rotados en la imagen derecha con respecto a la izquierda y viceversa. Por tanto, se trata de una ventaja importante en este tipo de imágenes y bajo la consideración del procesamiento de regiones, que debe explotarse naturalmente para establecer las correctas correspondencias.



Por el contrario, la invarianza a los cambios de escala no presenta una ventaja tan clara como en el caso de las rotaciones. En efecto, podría decirse que más bien suponen una cierta desventaja. Esto es así porque una región con una determinada dimensión en una imagen podría corresponderse con otra región de distinta dimensión, mayor o menor, en la otra imagen, ambas regiones con forma y aspecto similares, procedentes de distintos troncos. Evidentemente, esto podría conducir a errores en la correspondencia, por lo que se hace necesario aplicar ciertas restricciones relativas al tamaño de las regiones para evitar estos efectos colaterales no deseados, tal y como se verá posteriormente. No obstante, a pesar del inconveniente expresado anteriormente, también posee su ventaja derivada de la geometría del sistema. En efecto, volviendo sobre el análisis exhaustivo realizado en la sección 4.2.4 sobre las distintas proyecciones del mismo tronco 3D en las imágenes, según su ubicación en relación a los hemisferios, un tronco de árbol próximo a la posición donde se captura una de las imágenes se proyecta bajo un área mayor que la generada por su proyección en la imagen capturada con la cámara en una ubicación más alejada. La conjunción de la propiedad de invarianza a cambios de escala de los momentos de Hu y los tamaños de las áreas, permite abordar esta problemática satisfactoriamente.

Para superar los problemas mencionados y al mismo tiempo aprovechar toda la información disponible proporcionada por el dispositivo de medida y su geometría misma, se ha diseñado el procedimiento que se describe a continuación. Antes de su descripción, introducimos algunas definiciones necesarias para su comprensión.

Sean  $\{L_1, L_2, \dots, L_{NL}\}$  y  $\{R_1, R_2, \dots, R_{NR}\}$  dos tipos de descriptores de características representando las regiones segmentadas y etiquetadas en las imágenes izquierda y derecha, respectivamente, de un par estereoscópico proporcionado por el dispositivo de medición estereoscópico.  $NL$  y  $NR$  son el número de características en la imagen izquierda y derecha respectivamente.

Como se ha mencionado previamente, cada característica  $L_i$  contiene: el área ( $A_i$ ), el centroide  $(x_i, y_i)$  y un vector con los siete momentos invariantes de Hu  $\mathbf{h}_i \equiv \{\phi_1^i, \phi_2^i, \dots, \phi_7^i\} \in \mathcal{R}^7$ , obtenidos a partir de las ecuaciones (3.20) a (3.26), es decir  $L_i \equiv \{A_i, (x_i, y_i), \mathbf{h}_i\}$ . De modo similar, para  $R_j \equiv \{A_j, (x_j, y_j), \mathbf{h}_j\}$  donde como antes

$\mathbf{h}_j \equiv \{\phi_1^j, \phi_2^j, \dots, \phi_7^j\} \in \mathbb{R}^7$ . Debido a que los siete momentos toman valores en diferentes escalas, realizamos una normalización sobre ellas mediante una transformación lineal, según la ecuación (4.32), referida a cada momento de Hu de forma individual, para que los valores de cada momento individual queden restringidos al rango  $[0,1]$ :

$$\Phi_k^h = \frac{\phi_k^h - m_k}{M_k - m_k}; \quad h = i, j \text{ y } k = 1, 2, \dots, 7 \quad (4.32)$$

donde  $M_k = \max\{\phi_k^i, \phi_k^j\}$  y  $m_k = \min\{\phi_k^i, \phi_k^j\}$ ;  $\forall i, j$ .

A partir de los momentos normalizados, según (4.32), definimos el conjunto  $D_k^j$  de todas las distancias entre una componente dada  $\Phi_k^i$  y el resto de componentes  $\Phi_k^j$ ,  $j = 1, 2, \dots, NR$ , según la ecuación (4.33).

$$D_k^j = \{|\Phi_k^i - \Phi_k^j|, \quad j = 1, 2, \dots, NR, \quad k = 1, 2, \dots, 7\} \quad (4.33)$$

En lugar de calcular una distancia global entre dos características involucrando sus vectores  $\mathbf{h}_i$  y  $\mathbf{h}_j$ , hemos preferido calcular las distancias individuales  $D_k^j$ . Esto se justifica por el hecho de que cada momento contribuye normalmente de una forma diferente, como un atributo para la correspondencia entre características  $L_i$  y  $R_j$ , de modo que cuando se utilizan distancias globales entre  $\mathbf{h}_i$  y  $\mathbf{h}_j$ , tales como la Euclídea, las contribuciones individuales podrían enmascararse.

Una vez introducidas las definiciones anteriores, ahora el problema se convierte en el siguiente: dada una característica  $L_i$  en la imagen izquierda, ¿cual es su característica correspondiente  $R_j$  en la imagen derecha?. Una solución a este problema es la que se propone a continuación.

Siguiendo el trabajo de Scaramuzza y col. (2008) establecemos las siguientes condiciones, derivadas de los momentos invariantes de Hu, que deben cumplirse necesariamente. De este modo la decisión sobre si  $L_i$  y  $R_j$  se corresponden, a partir de  $\Phi_k^i$  y  $\Phi_k^j$ , se toma en base a las condiciones A a D siguientes:

- A.  $d_k^{ij} = \min\{D_k^i\} < T_1$ . Esta condición establece que la distancia mínima es, en efecto, la obtenida con  $\Phi_k^i$  y  $\Phi_k^j$  y además menor que un determinado umbral  $T_1$ , esto es, sólo se aceptan distancias por debajo de dicho umbral.  $T_1$  se fija a 0.3 tras diversos experimentos.
- B.  $d_k^{ij} = \min\{D_k^i\} < T_2 \langle D_k^i \rangle$ . Esta segunda condición se interpreta en el sentido de que dicha distancia mínima debe ser suficientemente menor que la media de las distancias de la región en cuestión con respecto al resto de distancias de regiones candidatas;  $\langle D_k^i \rangle$  representa el valor de dicha media, mientras  $T_2$  es un umbral cuyo rango de valores puede variar de 0 a 1. En este trabajo, ha sido fijado a 0.5 siguiendo las pautas establecidas en los trabajos de investigación de Scaramuzza y col. (2008).
- C. La diferencia entre  $d_k^{ij}$  y la segunda distancia mínima considerando el resto de regiones candidatas  $d_k^{ih} = \min\{D_k^i\}$  con  $h = 1, 2, \dots, NR$  y  $h \neq j$ , es menor que un determinado umbral  $T_3$ , fijado a 0.3 en nuestros experimentos. Esta condición se establece con el fin de garantizar que existe una cierta diferencia significativa entre la menor distancia y la siguiente.
- D. Las áreas  $A_i$  y  $A_j$  no difieren entre ellas más de un 33%, que se establece con el fin de evitar que áreas de muy diferente tamaño puedan corresponderse debido a la invarianza de escala de los momentos de Hu.

#### 4.3.4 Orden posicional: ángulos

Basándonos en la restricción de orden posicional, cabe esperar que las posiciones relativas entre dos regiones en una imagen hayan de preservarse en la otra imagen del par estereoscópico para las correspondencias que se suponen verdaderas. La aplicación de la restricción del orden posicional se limita a regiones con alturas y áreas similares en la misma imagen, estableciéndose además que ambas áreas deben sobrepasar un umbral  $T_4$  fijado a 6400 en este trabajo. La similitud se define del mismo modo que en el punto D, es decir, cuando la relación entre ambas áreas se sitúa por debajo del 33%. De este modo lo que se intenta evitar son posibles violaciones de esta restricción basándonos en relaciones de proximidad y lejanía de los troncos con respecto

al dispositivo de medida en la escena 3D. En la sección 5.7 se proporcionan abundantes detalles en relación a este hecho sobre el par de imágenes estereoscópicas allí mostrado, que por otra parte se trata de un comportamiento general para el conjunto de pares de imágenes analizado durante el presente trabajo de investigación.

Por tanto, la materialización de esta restricción se concreta como sigue. Dado un determinado orden relativo entre dos regiones (troncos) en una imagen, digamos la izquierda, bajo esta restricción, según las condiciones anteriormente establecidas, se asume que el mismo orden se preserva en la imagen derecha para los troncos homólogos y viceversa cuando se toma como referencia la imagen derecha, véase la figura 1.5(a,b). El orden posicional se establece en función de las posiciones relativas de los centroides en las imágenes, es decir, que en realidad esta restricción se aplica sobre los centroides representativos de las regiones bajo consideración, teniendo en cuenta sus posiciones espaciales en las imágenes en función de sus coordenadas polares, y por tanto debe verificarse en función de la orientación de los centroides en grados, según se expresa en la figura 4.1(b).

El siguiente ejemplo pedagógico permite realizar algunas aclaraciones sobre lo anteriormente expuesto. Con referencia a la figura 5.10(a) y (b), el tronco etiquetado como 2 en la imagen izquierda se corresponde con el etiquetado como 1 en la imagen derecha. La región etiquetada como 1 en la imagen izquierda se corresponde con la etiquetada como 5 en la derecha. Siguiendo el sentido de las agujas del reloj la región 1 se encuentra antes que la región 2 en la imagen izquierda y sus correspondencias preservan este orden, es decir 5 se encuentra antes que 1 en la imagen derecha. El criterio “se encuentra antes” se establece considerando el ángulo de orientación de los respectivos centroides en coordenadas polares.

#### 4.3.5 Resumen del proceso

Como colofón a lo expuesto previamente, el proceso de correspondencia completo puede resumirse como sigue:

##### *Correspondencia de Izquierda a Derecha*

Para cada región  $L_i$  en la imagen izquierda se buscan regiones candidatas  $R_j$  en la imagen derecha, de acuerdo con los pasos siguientes.

1. Aplicación de la restricción *epipolar*: sólo se consideran potenciales correspondencias de  $L_i$  aquellas  $R_j$  que cumplan la restricción de epipolaridad, tal y como se define en la sección 4.3.2. Tras la aplicación de dicha restricción,  $L_i$  posee como potenciales candidatos una lista  $l_i$  de  $n$  regiones en la imagen derecha, sea  $l_i \equiv L_i \rightarrow \{R_{j1}, \dots, R_{jn}\}$  dicha lista, donde  $j1, jn \in \{1, \dots, NR\}$ .
2. Aplicación de la condición D definida en la sección 4.3.3 a la lista  $l_i$ , excluyendo de  $l_i$  aquellos candidatos que no cumplan tal condición D.
3. Seguidamente se aplican las condiciones A, B y C definidas en la sección 4.3.3 sobre la actual lista  $l_i$ . Para cada par  $(L_i, R_{jn})$  obtenido de  $l_i$ , se determina si  $L_i$  y  $R_{jn}$  se corresponden basándose en el  $k$ -ésimo momento invariante de Hu según las mencionadas condiciones A, B y C. Se define  $lk$  como el número de posibles correspondencias individuales, así llamadas por haberse obtenido con el momento  $k$ -ésimo.

#### *Correspondencia de Derecha a Izquierda*

4. Para cada región  $R_j$  en la imagen derecha se buscan regiones candidatas  $L_j$  en la imagen izquierda, siguiendo los mismos pasos que en el caso anterior. De este modo, ahora se construye una lista de candidatos  $r_j$  y se obtiene un número  $rk$  de correspondencias individuales según cada uno de los siete momentos invariantes de Hu. Seguidamente se aplica la restricción *epipolar* siguiendo las mismas líneas epipolares que las utilizadas en el proceso de Izquierda a Derecha, si bien en sentido contrario.

#### *Decisión final: mayoría simple y unicidad*

5. Se decide que  $L_i$  se corresponde con  $R_j$ , si y sólo si  $lk + rk > U$ , donde  $U$  ha sido fijado a siete en nuestros experimentos. Este valor se ha determinado teniendo en cuenta que el máximo valor que la suma  $lk + rk$  puede alcanzar es catorce, por lo que un valor mayor que siete representa la mayoría.
6. Si la correspondencia entre  $L_i$  y  $R_j$  es inequívoca, el problema del emparejamiento entre ambas características queda resuelto; en caso contrario, se produce una situación de ambigüedad, donde la condición establecida en el

punto cinco se cumple para más de un par de correspondencias. Para resolver esta ambigüedad se aplica la restricción de *orden posicional*, donde las correspondencias inequívocas que ya han sido establecidas, permiten discernir en el caso de ciertas ambigüedades. Bajo esta decisión se incluye la aplicación simultánea de las restricciones de *orden posicional* y *unicidad*.

Tabla 4.3 Parámetros y umbrales implicados en el proceso de correspondencia basado en regiones

Fase	Parámetro/Umbral	Descripción
Restricción epipolar	$desp = 10$ $maxSep = 22$	Parámetros involucrados en la aplicación de la restricción epipolar al proceso de correspondencia basado en regiones, sección 4.3.2.
Restricción de semejanza	$T_1 = 0.3$ $T_2 = 0.5$ $T_3 = 0.3$ $difArea = 33$	Umbrales involucrados en la aplicación de la restricción de semejanza en el proceso de correspondencia basado en regiones, sección 4.3.3.
Restricción de orden posicional	$T_4 = 6400$	Umbral para decidir si se tiene en cuenta una región o no. Su área debe sobrepasar un umbral $T_4$ . Sección 4.3.4.
Restricción de unicidad	$U = 7$	Umbral para decidir si la región $L_i$ se corresponde con $R_j$ . Paso 5 de la sección 4.3.5.



## **Análisis de resultados**

### **5.1 Introducción**

A lo largo de este trabajo se han abordado dos tipos de estrategias orientadas a la resolución del problema de la correspondencia en visión estereoscópica en imágenes procedentes de dos tipos de bosques diferentes y además tomadas bajo muy distintas condiciones externas de iluminación, tal y como se expuso en el capítulo uno. Ello a pesar de que su captura se realiza con el mismo dispositivo de medición estereoscópica.

Por las razones allí expuestas y debido a la distinta naturaleza y características de las imágenes en los mencionados entornos forestales no estructurados, se han propuesto las dos estrategias mencionadas; la primera con fundamento en los píxeles y la segunda basada en las regiones. En ambos casos el objetivo final consiste en identificar en las dos imágenes del par estereoscópico la ubicación de los troncos, con el objetivo de proceder a realizar medidas de distancias con diferentes propósitos. Si bien, este trabajo se ha centrado exclusivamente en resolver el problema de la correspondencia estereoscópica, como paso previo e imprescindible del posterior proceso de medición.

Tanto en el enfoque basado en píxeles como en regiones, se han diseñado dos procesos para abordar el problema, estos son: segmentación y correspondencia propiamente dicha. A lo largo del presente capítulo se proporcionan los resultados



obtenidos en cada uno de ellos, distinguiéndose en cada caso entre la propuesta basada en píxeles y en regiones.

Dado que el sistema de medición estereoscópico utilizado es el mismo, a continuación proporcionamos algunos detalles relativos a su estructura y diseño. Seguidamente introducimos las pautas necesarias para seguir este capítulo en lo relativo a la obtención y análisis de resultados desde las dos perspectivas abordadas en este trabajo.

### 5.1.1 Características y operación del sistema de medición estereoscópico

La geometría del sistema se basa en el esquema gráfico descrito en la figura 4.3, donde la línea base de separación de ambas cámaras es de  $1m$ , que en dicha figura viene representada por la variable  $y_{12}$ . El fundamento de los procesos diseñados se basa en gran medida en la geometría del sistema estereoscópico, y más específicamente en las características del sistema óptico utilizado, mediante el cual se realiza la proyección de la escena tridimensional sobre la imagen bidimensional. En efecto, la cámara está equipada con una lente de las denominadas de “ojo de pez” Nikon® FC-E8, figura 5.1, caracterizadas principalmente por poseer un amplio campo de visión (*Field Of View*, *FOV*), siendo en el caso que nos ocupa de  $183^\circ$ , tal y como puede apreciarse en la representación de la figura 5.2(a) (Bakstein y Pajdla, 2002).



Figura 5.1 Lente de ojo de pez Nikon® FC-E8

En la figura 5.2(b) se muestra de forma esquemática la proyección de la escena tridimensional, a través de la lente del tipo “ojo de pez”, sobre el sensor CCD con el que está equipada la cámara digital, en este caso del modelo Nikon® Coolpix® 4500, que aparece en la figura 5.3. Se trata de un sensor común, que proporciona imágenes en las tres bandas típicas del espectro visible representadas en el modelo de color *RGB*. Dichas imágenes son del mismo tipo que las mostradas en los capítulos precedentes, cuyas dimensiones espaciales son  $M \times N$  en  $x$  e  $y$  respectivamente, si bien, desde el punto de vista del análisis sólo los píxeles pertenecientes al círculo central son los realmente válidos, que alcanzan el número de 6586205 píxeles.

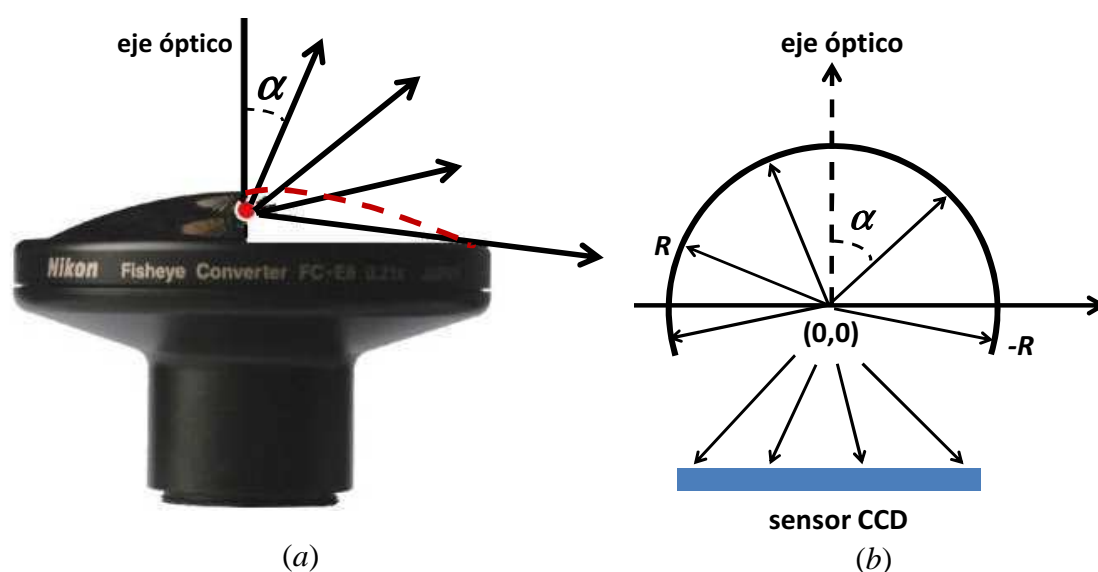


Figura 5.2 Descripción del FOV de la lente Nikon® FC-E8

El sistema estereoscópico en su conjunto es el que se muestra en la figura 5.4, diseñado específicamente para el CIFOR del INIA por la empresa “Consultoría Ferroviaria y Medioambiental, S.L.” ubicada en C/ Isaac Albéniz, 33, Las Rozas, 28290, Madrid. Este dispositivo se encuentra patentado con número de patente MU-200501738. Como puede observarse, se trata de un trípode sobre el que se sitúa un raíl, que permite el deslizamiento de la única cámara disponible, de suerte que en una posición de la cámara puede obtenerse una imagen y posteriormente, tras su deslizamiento por dicho raíl hasta una cierta distancia con respecto a la ubicación anterior, se obtiene una segunda imagen. En el caso de las imágenes analizadas en este trabajo, estos posicionamientos se corresponden con los extremos del raíl y el desplazamiento constituye la totalidad del raíl, que como se ha mencionado es de 1m y

constituye la línea base. Ambas imágenes constituyen el par estereoscópico para el análisis.



Figura 5.3 Vista de la lente Nikon® FC-E8 acoplada sobre la cámara Nikon® Coolpix® 4500

Por otra parte, como bien es sabido, cuanto menor sea la línea base, menor resulta también la disparidad de las imágenes en el par estereoscópico. En el caso de las parcelas bajo análisis, la densidad no es relativamente alta en comparación con otros sistemas forestales, por lo que se pueden utilizar mayores distancias de separación de la línea base sin que los solapamientos y oclusiones entre los troncos de los árboles lleguen a constituir un problema que impida llevar a cabo los análisis de las imágenes, como se ha podido comprobar a lo largo de los diferentes experimentos realizados. Dado que, hasta donde hemos podido investigar, no existen en la literatura estudios previos en este sentido, se ha optado por tomar las imágenes con la máxima separación de la línea base, obteniéndose valores de disparidad máximos de  $8^\circ$  de ángulo en las imágenes de pinares de pino silvestre (*Pinus sylvestris* L.). No ocurre lo mismo en las imágenes procedentes de los rebollares (*Quercus pyrenaica* Willd.), donde la densidad es mayor y las disparidades se incrementan considerablemente con respecto a las anteriores, llegándose a valores de  $22^\circ$  de disparidad. No obstante, debido a que en este caso el enfoque del análisis se basa en regiones, lo importante es conseguir aislar unas de otras en las dos imágenes, independientemente de su ubicación, para su posterior identificación. Los atributos para su identificación poseen la fortaleza suficiente como para ser independientes de dicha ubicación. En consecuencia, por las razones expuestas

anteriormente, a lo largo de los experimentos llevados a cabo en este trabajo de investigación se ha preferido tomar las imágenes con la línea base fijada a 1m. No obstante, la investigación sobre distintos valores de línea base se propone como tarea de futuro en el capítulo seis.

Obsérvese, a título de curiosidad, la varilla que aparece a la derecha del raíl en posición vertical. Ella es la que genera la marca en la parte inferior de la que hemos denominado imagen izquierda a lo largo de nuestro trabajo, que se proyecta en la imagen derecha prácticamente inapreciable sobre la misma posición inferior. En cualquier caso, aunque en la figura 5.4 aparece en la parte derecha del raíl, estas posiciones son relativas, ya que visto el sistema desde la parte opuesta, ésta se situaría en la izquierda, tal y como se ha considerado a lo largo de este trabajo.



Figura 5.4 Vista del sistema de medición estereoscópico general.

Por otro lado, para minimizar en lo posible el efecto causado por el hecho de tomar una imagen y tener que desplazar la cámara para la obtención de la segunda, el trípode está dotado de un sistema de anclaje en cada una de sus patas, que garantiza un posicionamiento lo más estable y robusto posible frente a este desplazamiento. Además, un tensor formado por una varilla rígida, uniendo un extremo del raíl con una de las patas del trípode permite mantener aquél en una posición lo más invariable posible a los efectos del movimiento. En cualquier caso, nada impide que en el futuro el sistema pueda evolucionar técnicamente, de suerte que se diseñe de forma que se equipe con dos

cámaras de disparo simultáneo para evitar desplazamientos indeseados del sistema entre capturas.

Como ya se ha mencionado en los capítulos precedentes, el objetivo último del sistema de visión estereoscópica consiste en obtener mediciones sobre los árboles que aparecen en las imágenes en relación a su ubicación y características en la parcela del bosque donde se capturan dichas imágenes (Montes y col. 2009).

El protocolo seguido durante la captura de las imágenes se realiza según se describe a continuación. Sobre el área boscosa se dispone de forma permanente de una serie de parcelas perfectamente identificadas geográficamente. Las parcelas de muestreo se separan entre sí de forma que entre los centros de las mismas las distancias varían entre 100 y 1000m. Dichas parcelas constituyen una muestra de la población bajo análisis, que es toda la masa arbolada, pudiendo estimarse parámetros de la población a partir de la muestra. El centro de las parcelas se ubica en una posición de coordenadas geográficas conocidas, obtenidas previamente vía GPS. En estas parcelas, para comparar los resultados, se han medido los árboles incluidos en un círculo de radio 25m. En este sentido, los árboles situados más allá de los 25m se proyectarán también sobre las imágenes, si bien con unas dimensiones relativamente pequeñas en comparación al resto, que permitirán fácilmente su exclusión durante el procesamiento de las imágenes. Este hecho se tiene en cuenta durante el diseño de las estrategias propuestas.

Por otra parte, como se menciona en el capítulo uno y se clarifica al hablar de la geometría del sistema en la sección 4.2.2, el sensor se posiciona bajo la dirección geográfica identificable, la primera imagen del par se obtiene situando la cámara en el extremo orientado al Sur de la guía y la segunda imagen se captura con la cámara situada en el extremo de la guía orientado al Norte, y ambas con línea base de 1m. Esto permite que diferentes mediciones espaciadas en el tiempo, sean obtenidas bajo los mismos criterios de medición, lo que permite comparar los valores de las variables medidas en fechas diferentes y así obtener crecimientos y evaluar los cambios que están teniendo lugar en la masa forestal.

En conjunto las pruebas de los métodos propuestos se han llevado a cabo con un conjunto de cuarenta pares de imágenes estereoscópicas. Se han utilizado veinte pares

para el procedimiento basado en píxeles y veinte para el procedimiento basado en regiones. Los pares de imágenes mostrados en este trabajo son un ejemplo representativo de ellas, caracterizándose por contener todos los aspectos relevantes derivados de la naturaleza misma de las imágenes y que se describieron en el capítulo uno.

A continuación, desde el punto de vista de las pruebas, se sigue el mismo planteamiento que en los capítulos anteriores, esto es, en primer lugar se analizan los resultados obtenidos mediante el procedimiento basado en píxeles y en segundo lugar los obtenidos mediante la estrategia basada en regiones. No obstante se separan los resultados para distinguir los obtenidos durante el proceso de segmentación de aquellos otros obtenidos mediante el proceso de correspondencia. De esta forma se pueden analizar y valorar los resultados de ambos procesos por separado, aunque bien es verdad que en el caso de la correspondencia se realizará un análisis global.

### 5.1.2 Enfoque basado en píxeles

Como ya se ha mencionado previamente, el objetivo se centra en los troncos de los árboles situados por debajo de los 25m en las parcelas objeto de análisis, de forma que para los píxeles analizados se computan los valores de disparidad en grados sexagesimales, según el procedimiento descrito en el capítulo cuatro.

Las pruebas han sido realizadas con veinte pares de imágenes estereoscópicas correspondientes a pinares, presentando todas ellas una alta variabilidad de los niveles de intensidad en sus tres componentes espectrales *RGB*, por haber sido captadas en días soleados y por tanto con alta variabilidad en la iluminación. Estos aspectos ya fueron puestos de manifiesto en el capítulo uno y matizados en el tres por su enorme relevancia, donde las figuras 1.3(a) y (b) muestran un par de imágenes estereoscópicas representativas de este tipo de imágenes. De los veinte pares disponibles, cuatro de ellos se han utilizado para el entrenamiento que conllevan los procesos de segmentación basados en clasificación de texturas mediante métodos de aprendizaje. En concreto, se trata de los siguientes, donde se identifica la naturaleza de los parámetros aprendidos:

- Los parámetros de los clasificadores *FC* y *PB* por separado: a) centros de las clases y grados de pertenencia para *FC* y b) centros de las clases y matrices de covarianza en el caso del *PB*.

- Se obtiene la relevancia de cada atributo a partir de los cuales se deducen las densidades *fuzzy* implicadas en la propuesta *CFI*.
- Se calcula la relevancia de cada atributo en la propuesta *SFI*, a partir de los cuales pueden obtenerse las densidades *fuzzy*, como en el caso anterior.
- Se calcula la relevancia de cada atributo en la propuesta *DES*, a partir de los cuales pueden obtenerse los vectores de medias  $\bar{v}_1$  y  $\bar{v}_2$ .
- Se obtiene la relevancia de cada atributo, a partir de las cuales pueden obtenerse los pesos *fuzzy*,  $w_1$  y  $w_2$ , implicados en la propuesta *MCDMF*.

En una segunda etapa, relativa a lo que en el capítulo tres se identificó como fase de decisión, obtenemos el mapa de disparidad inicial para cada par estereoscópico de los dieciséis pares restantes aplicando las propuestas siguientes píxel por píxel:

- *CFI*
- *SFI*
- *DES*
- *MCDMF*

A continuación, sobre cada mapa de disparidad obtenido mediante los procesos previos, se realiza un filtrado del mismo mediante los dos métodos siguientes:

- *FCM*
- *HNN*

La validación de las estrategias propuestas se lleva a cabo mediante el cómputo del error obtenido en cada método por comparación de las disparidades frente a las de referencia para cada par estereoscópico. El mapa de disparidades de referencia o “*ground truth*” se elabora contando con la colaboración de los expertos forestales, que tienen perfectamente identificados los árboles bajo análisis en cada parcela de muestra y por tanto en cada par de imágenes. De esta forma, dado el par de imágenes estereoscópicas a analizar, se toma como referencia la imagen izquierda, que es sobre la que se van a generar y computar las disparidades durante los procesos propuestos. A cada uno de los troncos identificados por los expertos, se le asigna manualmente un valor de disparidad delimitando su ubicación. Esta asignación se lleva a cabo tomando



la imagen izquierda del par estereoscópico y dejando sobre ella únicamente los troncos de interés, de forma que cada tronco quede aislado del resto. El resultado de esta operación se muestra en la imagen de la figura 5.5, donde se observan los troncos de interés extraídos a partir de la correspondiente imagen izquierda, en este caso para la imagen de la figura 1.3(a). A continuación se identifican estos troncos mediante el proceso de etiquetado automático de MATLAB (2010) basado en el procedimiento de etiquetado de componentes conexas descrito en Haralick y Shapiro (1992) y reproducido en Pajares y Cruz (2007a). El resultado de esta operación se muestra en la imagen de la figura 5.6. Una vez que a cada tronco se le asigna una etiqueta, se establece, también de forma manual, la relación entre cada etiqueta y la disparidad que le corresponde. Para la asignación de disparidades a cada uno de estos troncos, se procede identificando manual y visualmente al menos 50 puntos característicos sobre el tronco en cuestión, en todas sus partes tanto en la imagen izquierda como en la derecha, es decir, puntos en la base y a distintas alturas del tronco a lo largo de toda su extensión. Para cada punto obtenido en coordenadas cartesianas se computa la diferencia en ángulos en coordenadas polares, que es en realidad su disparidad. El valor medio de las disparidades manuales así obtenidas, es el que finalmente se asigna a la etiqueta que le corresponde a dicho tronco. De esta forma se obtienen los valores de disparidad promediados y mostrados en la tabla 5.1, para el par de imágenes que venimos utilizando a lo largo de este trabajo.



Figura 5.5 Aislamiento de los troncos de interés



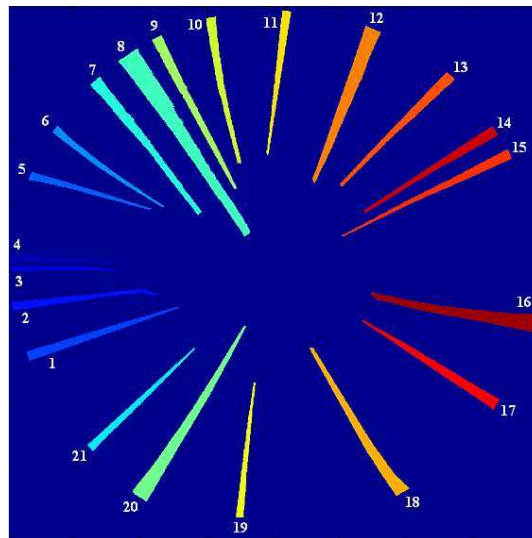


Figura 5.6 Etiquetado de los troncos

Tabla 5.1 Disparidades promediadas para el *ground truth*

Ground truth	
Región	Disparidad promedio
1	4.55
2	3.92
3	2.44
4	3.14
5	3.32
6	3.77
7	4.88
8	6.93
9	3.33
10	1.80
11	0.26
12	2.21
13	3.02
14	2.79
15	4.74
16	5.62
17	3.00
18	2.09
19	0.68
20	4.69
21	2.48

Los porcentajes de error para la valoración de los métodos de correspondencia propuestos en este trabajo se computan como sigue. Cuando la disparidad de un píxel dado, obtenida por cualquiera de los procesos propuestos, y la asignada manualmente no difiere mas de  $0.5^\circ$ , que es la tolerancia considerada como aceptable por los expertos y válida para el análisis estadístico realizado en las parcelas, se considera que se ha producido un acierto, contabilizándose como tal. En caso contrario, se trata de un error.

La valoración sobre los porcentajes de acierto y error se realiza exclusivamente sobre los troncos de interés. Es claro que los valores de disparidad obtenidos para los píxeles situados fuera de los troncos, no ofrecen garantías suficientes de fiabilidad por las razones expuestas en los capítulos precedentes, principalmente en relación a los problemas derivados de la alta variabilidad en las intensidades de píxeles vecinos.

Para cada uno de los dieciséis pares de imágenes estereoscópicas usados en la fase de decisión y de los que se ha obtenido su mapa de disparidad, calculamos el error en la disparidad para los píxeles pertenecientes a los troncos y promediamos esos errores entre los dieciséis pares de imágenes. Como se ha mencionado previamente, el par de imágenes estereoscópicas mostrado en las figuras 1.3(a) y (b) es uno de los dieciséis pares utilizados en la fase de decisión para establecer las correspondencias. Sobre este par nos basamos utilizándolo como ejemplo para definir el proceso relativo al cómputo del error, el cual consta de los siguientes pasos:

- 1) Para cada par de imágenes estereoscópicas analizadas, designado como par  $i$ , se crea una variable  $E_i$ , que sirve para acumular los valores de error píxel a píxel.
- 2) Teniendo en cuenta el mapa de disparidad obtenido para un par estereoscópico por cualquiera de los métodos propuestos y el correspondiente *ground truth* para dicho par, se evalúa la diferencia de disparidad de los píxeles en el mapa en relación al mismo píxel en el *ground truth*. Si dicha diferencia resulta ser superior, en valor absoluto, al límite de tolerancia prefijado en  $0.5^\circ$ , el valor de disparidad para dicho píxel se contabiliza como error, en cuyo caso, se incrementa en uno el valor de la variable  $E_i$ , esto es  $E_i = E_i + 1$ . El porcentaje de error para el par de imágenes bajo análisis se contabiliza dividiendo el valor final de  $E_i$  por el número de píxeles etiquetados en su correspondiente *ground truth*, esto es, el número de píxeles que aparecen coloreados en la imagen de la figura 5.6.
- 3) Este proceso se repite para cada uno de los pares de imágenes analizadas, de suerte que al final se computa el valor medio del error sobre el total de

los pares, esto es  $\bar{E} = \frac{1}{N} \sum_{i=1}^N E_i$ , con  $N$  igual a dieciséis en este trabajo.

Estos valores promediados de error son los que aparecen posteriormente en las tablas de resultados.

En la sección 5.2 de este capítulo se presentan los resultados, referidos al enfoque basado en píxeles, como consecuencia de la aplicación del proceso de segmentación descrito en la sección 3.2, incluyendo las fases de entrenamiento y decisión. En la sección 5.4 se muestran los resultados obtenidos en el proceso de correspondencia descrito en la sección 4.2. Se realiza un análisis comparativo entre los resultados obtenidos por cada uno de los métodos objeto de investigación, *CFI*, *SFI*, *DES* y *MCDMF*. Dado que dichos métodos en realidad toman sus decisiones de acuerdo con una serie de criterios, descritos en la sección 3.2.4, mediante combinación de los mismos, se ha creído conveniente realizar la comparación entre los métodos que combinan estos criterios frente a los resultados obtenidos cuando se aplica cada criterio de forma individual. Esto nos permite identificar, entre otras cosas, la relevancia de cada criterio individual y las ventajas de la combinación. Por otra parte, al tratar los métodos combinados, se ha elegido la estrategia de combinación clásica propuesta por Yager (1988) para valorar frente a ella los resultados obtenidos por las propuestas de este trabajo, tal y como se describe más adelante.

Basándonos en los resultados de la comparación, se elige la mejor de todas las estrategias en función del menor porcentaje de error. Ésta es la que finalmente se utiliza como punto de partida para la mejora del mapa de disparidad mediante las estrategias de filtrado propuestas, a saber: *FCM* y *HNN*, cuyos resultados se presentan en la sección 5.5. Finalmente, se vuelven a comparar los resultados obtenidos frente a las imágenes de referencia o *ground truth* creadas a tal efecto, permitiendo de este modo realizar un resumen y valoración de resultados en la sección 5.6.

### 5.1.3 Enfoque basado en regiones

En este caso, las pruebas se han realizado con veinte pares de imágenes estereoscópicas procedentes de bosques de robles, habiendo sido obtenidas bajo las condiciones de escasa iluminación, tal y como se describe en el capítulo uno, donde aparece un par representativo de este conjunto de imágenes en las figuras 1.5(a) y (b).

En este enfoque utilizamos cuatro pares de ellas para el entrenamiento requerido, con el fin de obtener los valores de los umbrales  $U_1$ ,  $U_2$  y  $U_3$  implicados en el proceso de segmentación, sección 3.3, así como los umbrales  $T_1$ ,  $T_2$ ,  $T_3$  y  $T_4$  involucrados en el proceso de correspondencia, sección 4.3.

Una vez tenemos calculados estos valores, para cada uno de dieciséis pares de imágenes estéreo restantes realizamos el proceso completo mediante los procedimientos de segmentación y correspondencia estereoscópica descritos respectivamente en los capítulos tres y cuatro.

En este caso, al tratarse de correspondencia por regiones el cómputo de los porcentajes de error difiere del caso anterior. En efecto, puesto que las regiones que identifican los troncos están perfectamente etiquetadas a través del proceso de segmentación en sendas imágenes del par estereoscópico, resulta relativamente fácil, mediante simple inspección visual, determinar si el emparejamiento de regiones pertenecientes a las imágenes izquierda y derecha ha sido correcto o no. De este modo se obtienen los errores de correspondencia, realizando un promedio entre los dieciséis pares de imágenes disponibles para este propósito. El par de imágenes estereoscópicas mostrado en las figuras 1.5(a) y (b) corresponde a uno de los dieciséis pares utilizados en el proceso completo de correspondencia. En este capítulo se presentan los resultados obtenidos con este par de imágenes tanto en la etapa de segmentación como de correspondencia, para clarificar el tipo de resultados obtenidos bajo este enfoque. Si bien los mismos son perfectamente extrapolables al conjunto de pares de imágenes analizadas, sin que los restantes pares aporte ningún elemento relevante que no aparezca reflejado en el análisis sobre el par de referencia.

En la sección 5.3 se muestran los resultados obtenidos mediante el proceso de segmentación basado en regiones y descrito en la sección 3.3. En la sección 5.7 se muestran los resultados obtenidos mediante el proceso de correspondencia descrito en la sección 4.3. Por último, en la sección 5.8 se realiza un análisis y resumen de los resultados obtenidos en el proceso completo de correspondencia bajo el enfoque basado en regiones.

## 5.2 Resultados de la Segmentación basada en píxeles

A continuación se muestran los resultados relativos a la segmentación de los pares de imágenes estereoscópicas analizados. Como se ha mencionado previamente, se dispone de un conjunto de cuatro pares de imágenes con tal propósito, que se utilizan tanto para los procesos de entrenamiento involucrados en esta fase de segmentación como para el ajuste y obtención de otros parámetros necesarios en la fase de decisión, como se verá posteriormente. Pues bien, la primera fase es la de entrenamiento, de suerte que a partir del conjunto de pares de imágenes estereoscópicas disponibles para este propósito, seleccionamos manualmente muestras pertenecientes a las texturas de hierba y cielo, obteniendo un conjunto de 2550 muestras correspondientes a ambas texturas pertenecientes tanto a las imágenes izquierda y derecha de los pares estereoscópicos utilizados en esta fase de entrenamiento. Como puede observarse, en la imagen de la figura 1.3(a) o equivalentemente en la de la figura 1.3(b) y por extensión al resto de las imágenes de la misma naturaleza, las texturas pertenecientes a la hierba presenta valores de intensidad variable dependiendo de si los píxeles corresponden a un área soleada o situada en la sombra en el momento de la captura de la imagen. Circunstancia ésta bien conocida en el tratamiento de imágenes ya que en la formación de las imágenes interviene la componente de iluminación junto con la reflectancia, en este caso de la hierba (Pajares y Cruz, 2007a). Por lo tanto, para evitar problemas con los valores absolutos de las componentes espectrales  $R$ ,  $G$ ,  $B$ , y en particular en lo referente al tema de las texturas que caracterizan la hierba, procedemos a realizar una normalización de sus valores al rango  $[0,1]$ . Así, dada una muestra  $\mathbf{x} = (R, G, B)$ , ésta se normaliza como  $\mathbf{x} = (R/U, G/U, B/U)$  con  $U = R + G + B$ .

### 5.2.1 Agrupamiento Borroso (FC)

Con la normalización anterior los centros representativos de cada clase se obtienen mediante el proceso de entrenamiento descrito en la sección 3.2.3.2a a través de las ecuaciones (3.4) y (3.5) correspondientes al clasificador  $FC$ , cuyos valores son los mostrados en la tabla 5.2. El entrenamiento de este clasificador requiere el ajuste de varios parámetros involucrados en las ecuaciones anteriores, a saber: el peso exponencial  $e$ , que aparece en la ecuación (3.4); el umbral de convergencia  $\varepsilon$  y el máximo número de iteraciones. Para determinar estos valores se ha tomado como

referencia los experimentos llevados a cabo en Guijarro (2009), ya que en este trabajo se obtienen dichos parámetros a partir de dieciséis tipos de datos, de los cuales nueve proceden de la famosa base de datos “*Machine Learning Repository*” (Asuncion y Newman, 2010) y los siete restantes se reparten entre tres tipos de datos sintéticos y cuatro conteniendo texturas naturales de exterior, entre las que se incluyen dos de ellas similares a las utilizadas en el presente trabajo. En total se utilizaron 37160 muestras, de las cuales 5470 fueron aportadas por imágenes con texturas naturales. A este conjunto se han añadido ahora las 2550 muestras procedentes de los cuatro pares destinados a tal fin en el presente trabajo de investigación, lo que hace un total de 39710 muestras utilizadas para el entrenamiento. Se trata pues de un conjunto de muestras suficientemente amplio y general en el sentido de que no sólo contiene datos procedentes de imágenes sino de otras variadas fuentes, lo cual permite garantizar una cierta validez para los parámetros estimados, que finalmente resultan ser los siguientes:  $e = 2.1$ ,  $\varepsilon = 0.01$  y  $k_{max} = 50$ . Con los parámetros fijados a dichos valores, finalmente los centros de las clases para las texturas bajo el proceso de identificación son los proporcionados en la tabla 5.2.

Tabla 5.2 Centros de las clases para las texturas hierba y cielo

Cielo	Hierba
$\nu_1 = (0.18, 0.35, 0.48)$	$\nu_2 = (0.32, 0.43, 0.17)$

Como se ha descrito antes, el proceso de segmentación se aplica tanto a las imágenes izquierdas como derechas de los pares estereoscópicos con el fin de realizar los procesos de correspondencia de izquierda a derecha y de derecha a izquierda, descartando los píxeles pertenecientes a las texturas carentes de interés para la correspondencia. Obviamente, la identificación de estas texturas se lleva a cabo mediante el correspondiente proceso de entrenamiento.

La figura 5.7(a) y (b) muestra el resultado de aplicar a las imágenes del par estereoscópico de la figura 1.3(a) y (b) respectivamente el proceso de segmentación diseñado con tal propósito, es decir, identificando las texturas correspondientes a las hojas de los pinos mediante el proceso descrito en la sección 3.2.2 a través del cómputo de la varianza local; así como las texturas del cielo y la hierba que se identifican por medio del proceso de decisión basado en el clasificador  $FC$ , descrito en esa misma sección y utilizando los centros de las clases dados en la tabla 5.2.

Como fácilmente puede deducirse, las zonas etiquetadas en blanco identifican los píxeles pertenecientes a uno de los tres tipos de texturas que quedan fuera del objeto de interés, es decir, píxeles de hojas y zonas de texturas pertenecientes al cielo y la hierba sobre el suelo. Por el contrario, las zonas etiquetadas en negro, dentro del círculo de definición de la imagen, son los píxeles a ser procesados para tratar de encontrar sus correspondencias en la otra imagen del par estereoscópico. Entre ellos, obviamente, se encuentran los píxeles pertenecientes a los troncos, que constituyen nuestro objeto de interés. Como puede observarse la mayoría de los troncos aparecen en color negro. Existen algunos troncos, etiquetados como blancos, que representan los árboles situados muy lejos del sistema estereoscópico, concretamente fuera del área de los 25m de distancia, que se establece como la distancia umbral para el análisis. Dichos troncos no se consideran porque se sitúan fuera del interés relativo al análisis a realizar en la zona marcada a tal efecto.

En resumen, como puede observarse en la figura 5.7, los píxeles blancos han sido identificados como pertenecientes a uno de los tres tipos de texturas que no son de interés, es decir, han sido descartados para la fase de correspondencia, según el objetivo previsto. No obstante, y en relación a este objetivo, cabe apreciar sobre dichas imágenes algunos errores de clasificación en este tipo de texturas, lo que viene a corroborar la necesidad de ahondar más en la investigación en esta línea.

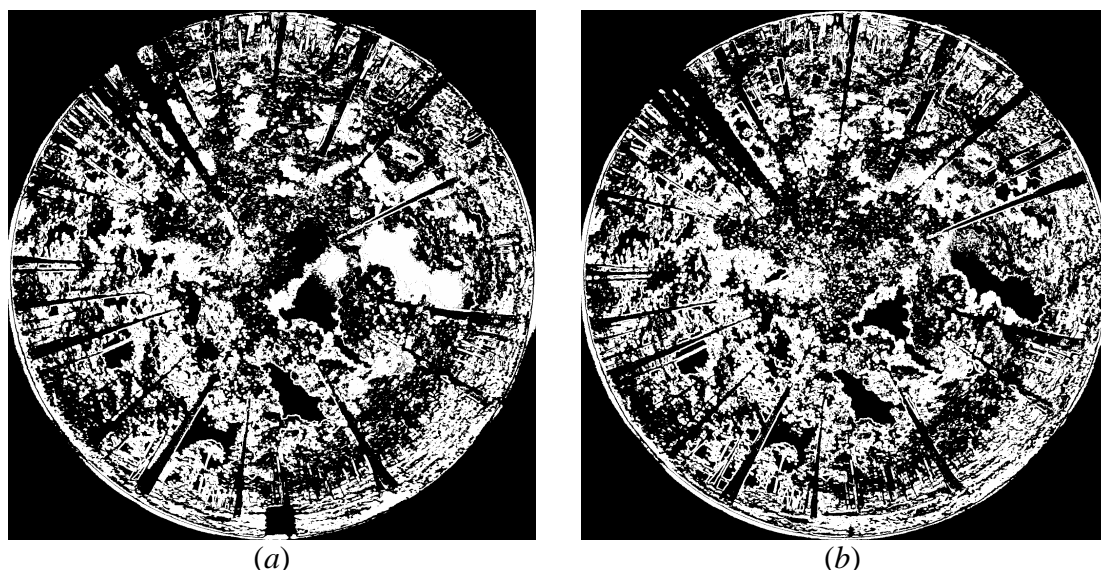


Figura 5.7 Imágenes izquierda y derecha segmentadas utilizando el descriptor de texturas junto con el clasificador *FC*.

### 5.2.2 Estimador Paramétrico Bayesiano (PB)

Durante la fase de entrenamiento correspondiente al clasificador *PB*, se obtienen los parámetros estimados en el correspondiente proceso, es decir, los centros de las clases hierba y cielo, tabla 5.3, así como las matrices de covarianza asociadas a dichas clases y proporcionadas en la tabla 5.4.

Tabla 5.3 Centros de las clases para las texturas hierba y cielo obtenidas mediante *PB*

Cielo	Hierba
$\mathbf{v}_1 = (0.16, 0.32, 0.52)$	$\mathbf{v}_2 = (0.31, 0.48, 0.14)$

Tabla 5.4 Matrices de covarianza de las clases para las texturas hierba y cielo obtenidas mediante *PB*

Cielo	Hierba
$\begin{bmatrix} 1.34 & 1.10 & 2.21 \\ 1.10 & 1.95 & 2.01 \\ 2.21 & 2.01 & 3.14 \end{bmatrix}$	$\begin{bmatrix} 0.78 & 0.56 & 0.83 \\ 0.56 & 1.23 & 1.01 \\ 0.83 & 1.01 & 0.65 \end{bmatrix}$

Las figuras 5.8(a) y (b) muestran el resultado de aplicar el proceso de segmentación al par estereoscópico representado en las imágenes de la figura 1.3(a) y (b) respectivamente involucrando al estimador paramétrico de Bayes, una vez que los parámetros asociados a este clasificador han sido estimados convenientemente. Como en el caso del *FC*, las texturas de las hojas se identifican mediante el proceso descrito en la sección 3.2.2 a través de la varianza, mientras que ahora las texturas del cielo y la hierba se identifican por medio del proceso de decisión basado en el clasificador *PB*, descrito en esa misma sección.

Si se comparan los resultados obtenidos por el clasificador *PB* frente a los obtenidos con *FC*, se observa claramente a partir de la observación de las figuras 5.8 y 5.7 respectivamente, cómo mediante *PB* se ha conseguido un incremento en el número de píxeles que identifican texturas cuyo objetivo es su exclusión del proceso de correspondencia. Esto, que a priori podría suponer un beneficio por la clara ventaja que su eliminación supone para el proceso de correspondencia, no lo es tal debido a que se ha producido lo que podríamos llamar un exceso de segmentación o sobre-segmentación, en el sentido de que ahora aparecen troncos de árboles y por tanto texturas de interés excluidas, lo cual es no deseable, ya que determinados troncos no intervendrían en el proceso. Esto ocurre principalmente en el caso de algunos troncos situados donde los rayos del sol inciden directamente sobre la escena proyectada en la



imagen, tal y como ocurre en el caso de la mitad derecha de las imágenes y más específicamente en el cuadrante superior derecho. En realidad este fenómeno se presenta porque las texturas de los árboles iluminados tienden a saturarse en intensidad, aproximándose al blanco en el modelo de color *RGB*, de aquí que el clasificador *PB*, mucho más sensible que el *FC* a este tipo de texturas iluminadas, las identifique como texturas de cielo.

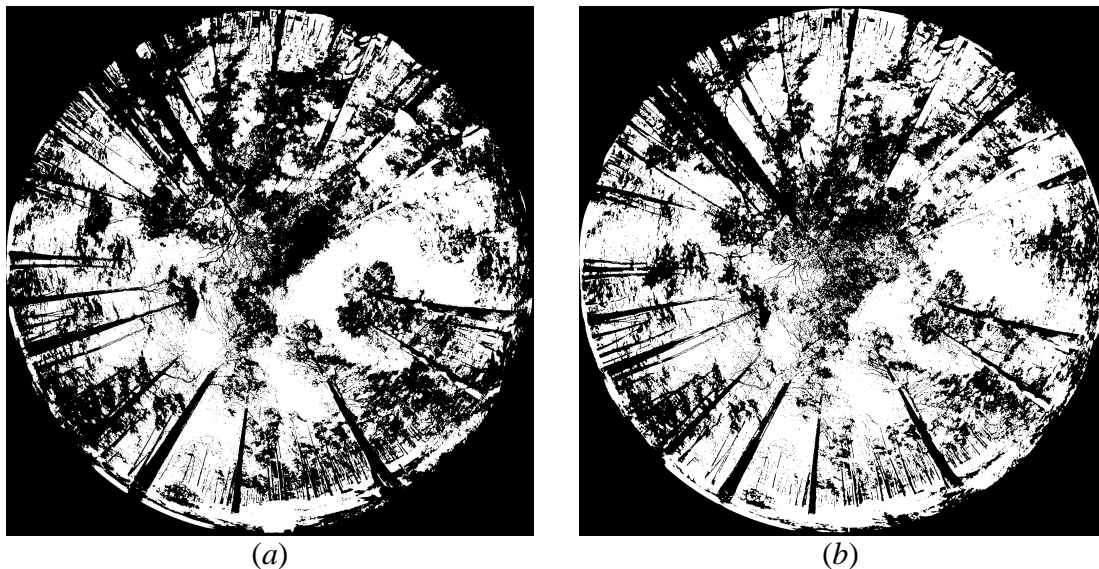


Figura 5.8 Imágenes izquierda y derecha segmentadas utilizando el descriptor de texturas junto con el clasificador *PB*.

### 5.2.3 Combinación Agrupamiento Borroso y Estimador Paramétrico Bayesiano

De las observaciones previas se deduce que es necesario el diseño de un clasificador que se sitúe en un término medio entre ambos, es decir, no sea tan pesimista como el *FC* en cuanto a la clasificación de las texturas objeto de exclusión ni tan optimista como el *PB*. Esta es una de las razones que justifican la utilización de una estrategia de clasificación combinada basada en los dos clasificadores mencionados, tal como se ha propuesto en la sección 3.2.3.3c. De aquí surge una nueva línea de investigación cuyos estudios previos se exponen en el apéndice B, relativos al diseño de una estrategia con el mayor grado de validez posible y a la vez para el mayor número de imágenes procedentes de entornos forestales.

Las figuras 5.9(a) y (b) muestran los resultados de la segmentación obtenidos mediante la utilización de la estrategia combinada propuesta en dicha sección, aplicada

al par de imágenes estereoscópicas de la figura 1.3(a) y (b), respectivamente. Naturalmente, se ha utilizado la identificación de las texturas correspondientes a las hojas de los pinos mediante la estrategia basada en la varianza.

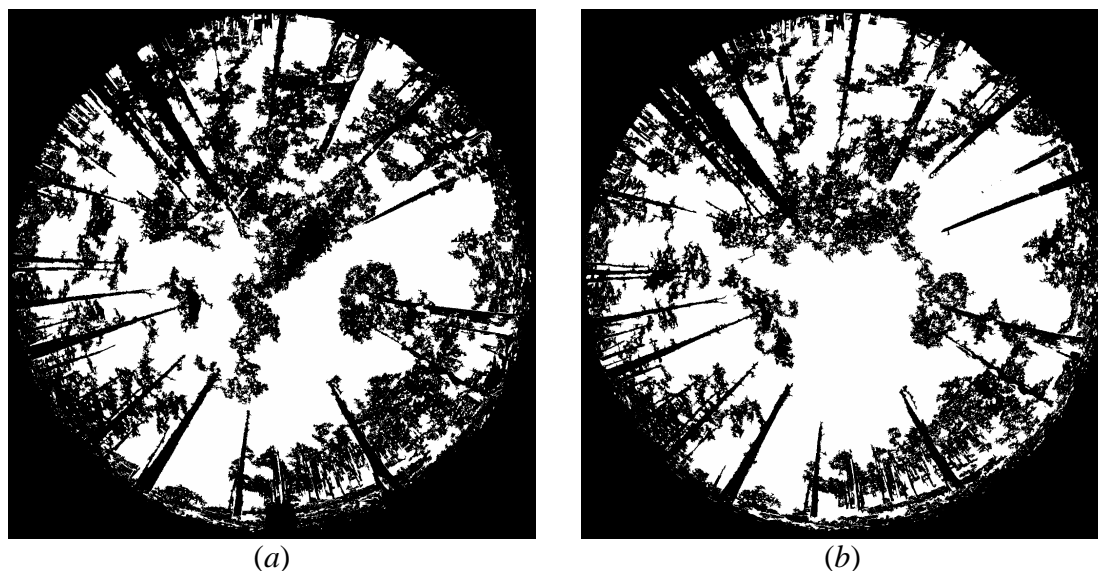


Figura 5.9 Imagen segmentada mediante la combinación de los clasificadores *FC* y *PB*

A la vista de los resultados mostrados en las imágenes de la figura 5.9, se observa un importante aumento del número de píxeles pertenecientes a las texturas a excluir en relación al resultado obtenido mediante el clasificador *FC*, a la vez que se resuelve el problema de los troncos fuertemente iluminados que aparecían como principal inconveniente en el caso del clasificador *PB*. En consecuencia, aunque persisten todavía texturas que no constituyen el interés del proceso de correspondencia, como es el caso de los árboles situados en la lejanía o las copas de los árboles, es evidente que la combinación de clasificadores puede considerarse como una estrategia digna de ser tomada en cuenta. A la vista de estos resultados, que son extensibles al resto de pares de imágenes analizadas, queda patente por una parte, la mejora de los resultados obtenidos mediante la combinación de clasificadores desde el punto de vista de su orientación hacia la correspondencia, y por otra parte, la justificación del uso de este enfoque combinado, que a su vez abre nuevas perspectivas de investigación en esta línea.

Conviene recalcar el hecho de que las anteriores apreciaciones son igualmente observables en el resto de los pares de imágenes estereoscópicas analizadas, tanto en lo que se refiere a la utilización de los clasificadores individuales *FC* y *PB* como en su versión combinada.

### 5.3 Resultados de la Segmentación basada en regiones

En primer lugar se utilizan cuatro pares de imágenes disponibles para este propósito con el fin de obtener los parámetros  $U_1$ ,  $U_2$  y  $U_3$  involucrados en este enfoque basado en regiones. Tras diversas pruebas de ensayo y error, estos parámetros son fijados a 10, 0.3 y 120 respectivamente tal y como se expresa en la sección 3.3.2, tras lo cual se está en disposición de pasar a la siguiente fase de decisión.

Las figuras 5.10(a) y (b) muestran las regiones extraídas mediante el proceso de segmentación descrito en la sección 3.3. Tras dicho proceso, cada región aparece marcada con una única etiqueta numérica, cuyo valor se muestra próximo a la región que representa sobre las respectivas imágenes izquierda y derecha de dichas figuras. Para una mejor visualización de estos resultados, se ha representado cada una de estas imágenes con sus regiones mostradas en color, donde cada color representa precisamente esa etiqueta, variando desde la tonalidad azul para la etiqueta 1 hasta la roja ocre de la 20 en la imagen izquierda o la 19 en la derecha, pasando por toda la gama de colores intermedios. Al fondo de la imagen, que evidentemente no representa región específica alguna, se le ha asignado el valor cero. En la parte inferior de ambas imágenes se muestra la escala de color correspondiente, según los valores numéricos de dichas etiquetas.

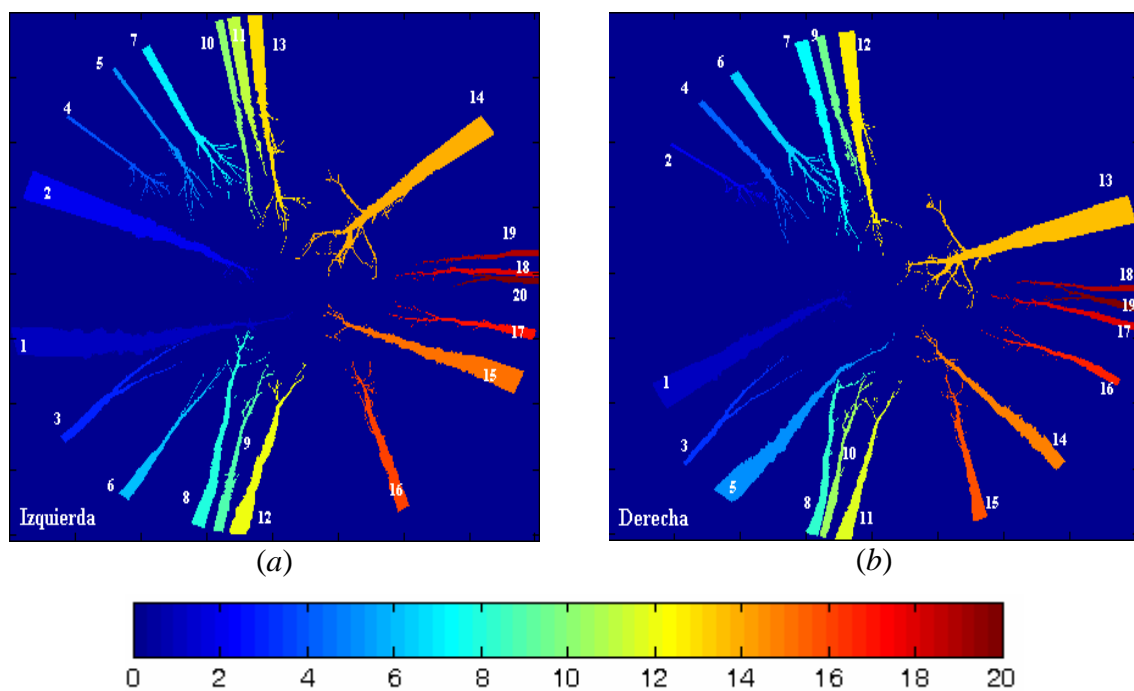


Figura 5.10. Etiquetado de regiones: (a) imagen izquierda, (b) imagen derecha. Cada región aparece identificada por un único número.

A partir de los resultados mostrados en las imágenes de la figura 5.10, se pueden deducir las siguientes conclusiones sobre el proceso de segmentación aplicado sobre ellas:

1. Las regiones han sido completamente separadas, incluso a pesar de la existencia de regiones muy próximas entre sí, lo que ha permitido su etiquetado de manera independiente. Esto ocurre con las regiones marcadas como 10 y 11 ó 18 y 20 en la imagen izquierda, así como con las regiones 8 y 10 en la imagen derecha. Gracias a este proceso de separación, dichas regiones se asemejan a los troncos de cada árbol de forma independiente, tal y como es deseable. De otro modo los pares de regiones anteriormente mencionadas hubiesen aparecido fusionadas como una única región, lo que hubiera impedido la extracción de propiedades relativas a la identificación de un único tronco representando a un árbol concreto.
2. Además, el procedimiento de segmentación es capaz de extraer regiones correspondientes a troncos, que están situados relativamente lejos del sensor, es decir, fuera del área de la parcela de muestra que constituye el área de interés. Esto es exactamente lo que ocurre con las regiones etiquetadas como 4, 5, 18, 19 y 20 en la imagen izquierda y 2, 17, 18 y 19 en la imagen derecha. A pesar de que tales regiones están fuera de nuestro interés, bajo el punto de vista del límite impuesto por las dimensiones de la parcela de muestra, con radio inferior a  $25m$ , hemos preferido incluirlas en el proceso de correspondencia, porque en un futuro es posible que el área bajo análisis pudiera ampliarse. Además de esto, la inclusión de estas imágenes permite verificar la robustez del proceso de correspondencia, dada su menor área relativa respecto a la de los troncos de la parcela. En cualquier caso, su exclusión es una tarea sencilla porque todas ellas cumplen que sus áreas poseen un valor por debajo de 6400 píxeles, que es precisamente el umbral  $T_4$  aplicado para la restricción de orden posicional descrita en la sección 4.3.4.
3. Finalmente, mediante las operaciones morfológicas aplicadas, el proceso de segmentación es capaz de romper ciertos enlaces existentes entre determinadas regiones, permitiendo su mejor asociación con los troncos reales. Esto es lo que ocurre entre las regiones 5 y 8 en la imagen derecha, donde aparecen dos ramas

en la parte superior de los árboles que se presentan superpuestas. Sin esta ruptura, ambas regiones se etiquetarían como una única región y su correspondencia con las regiones correspondientes en la imagen izquierda, que se encuentran perfectamente separadas, no sería posible, además de no corresponderse con árboles individuales, como se ha mencionado previamente.

Las anteriores apreciaciones son completamente aplicables al resto de pares de imágenes analizadas mediante esta estrategia basada en regiones, tal y como se ha podido constatar durante su estudio.

## 5.4 Resultados de la Correspondencia basada en píxeles

Tras la segmentación basada en píxeles para las imágenes de pinos, el siguiente paso consiste en tratar de establecer la correspondencia de los píxeles que no han sido identificados como pertenecientes a los tres tipos de texturas descritas. Este proceso se lleva a cabo mediante las técnicas de correspondencia descritas en el capítulo cuatro: *CFI*, *SFI*, *DES* y *MCDMF*. Tal y como se expuso en dicho capítulo, cada uno de los métodos requiere para su aplicación el cómputo previo de una serie de parámetros; así *CFI* y *SFI* necesitan las relevancias de cada atributo interviniente; *DES* los vectores de medias de las clases involucradas y *MCDMF* los pesos asociados a cada criterio. Pues bien, para cada uno de dichos métodos se exponen y analizan, en primer lugar, los resultados obtenidos en relación a sus parámetros, que se obtienen a partir de los cuatro pares de imágenes estereoscópicas destinados a tal efecto. En segundo lugar se proporcionan los detalles más significativos sobre los resultados obtenidos por cada método tras su aplicación sobre el conjunto de los dieciséis pares de imágenes de prueba disponibles.

### 5.4.1 Integral Fuzzy de Choquet (CFI)

#### 5.4.1.1 Cálculo de la relevancia para cada atributo

Tal y como se ha descrito en la sección 4.2.4.1, este proceso requiere, como paso previo a su aplicación, el cómputo de las densidades *fuzzy* asociadas con cada uno de los seis atributos que participan en el proceso, y a través de ellas se obtienen las relevancias asociadas a cada criterio. Sobre los cuatro pares de imágenes estéreo utilizadas para este propósito, se aplica el proceso de segmentación basado en píxeles. Tras el proceso de

segmentación, se toma cada uno de los seis atributos por separado, a saber: *a*) coeficiente de correlación, *b*) color, *c*) textura, *d*) módulo del gradiente, *e*) ángulo del gradiente y *f*) Laplaciana. Así, dado un píxel en una imagen, se obtienen medidas de similitud, según la ecuación (4.5), con respecto a sus candidatos en la otra imagen del par estereoscópico para cada atributo por separado. De esta forma se obtienen seis mapas de disparidad generados respectivamente por cada uno de los seis atributos basándonos en el criterio de máxima similitud. La correspondencia se establece siguiendo el mismo protocolo definido en el capítulo cuatro, esto es, de izquierda a derecha y de derecha a izquierda. Para cada mapa se calcula el correspondiente porcentaje de error con respecto a su correspondiente *ground truth*. Los porcentajes de error para los cuatro pares estereoscópicos se promedian obteniendo finalmente los valores,  $p_h$ , mostrados en la tabla 5.5, donde  $p_a, p_b, p_c, p_d, p_e, p_f$  se refieren respectivamente a los atributos de correlación, color, textura, módulo del gradiente, ángulo del gradiente y Laplaciana. De este modo, las relevancias individuales asociadas a cada atributo se calculan según la siguiente expresión:  $g^h = (100 - p_h) / \sum_k (100 - p_k)$ ,  $h, k = a, b, c, d, e, f$ , cuyos valores son los mostrados en la tabla 5.5 para cada atributo individual.

Tabla 5.5 Porcentajes promediados de error y relevancias individuales obtenidas para cada atributo con el conjunto de pares de imágenes de entrenamiento.

Porcentajes promediados de error $p_h$ y relevancias de los atributos $g^h$		
Criterio	$p_h$	$g^h$
$s_a$	28	0.150
$s_b$	10	0.187
$s_c$	14	0.178
$s_d$	9	0.189
$s_e$	30	0.145
$s_f$	27	0.151

Para una mayor claridad, las figuras 5.11 y 5.12 muestran gráficamente, los porcentajes promediados de error y las relevancias individuales de la tabla 5.5 para cada uno de los seis atributos.

Se observa a partir de los valores obtenidos que el atributo más relevante se corresponde con el módulo del gradiente. Nótese que a partir de las relevancias individuales, se está en disposición de calcular las densidades *fuzzy* en cada proceso de decisión asociado a cada píxel, según la ecuación (4.7).

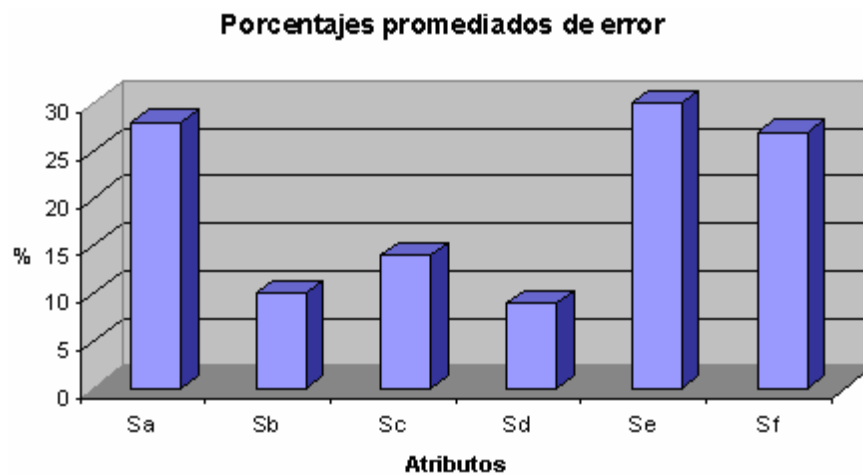


Figura 5.11 Porcentajes promediados de error obtenidos para cada atributo individual a partir del conjunto de pares de imágenes estereoscópicas de entrenamiento.

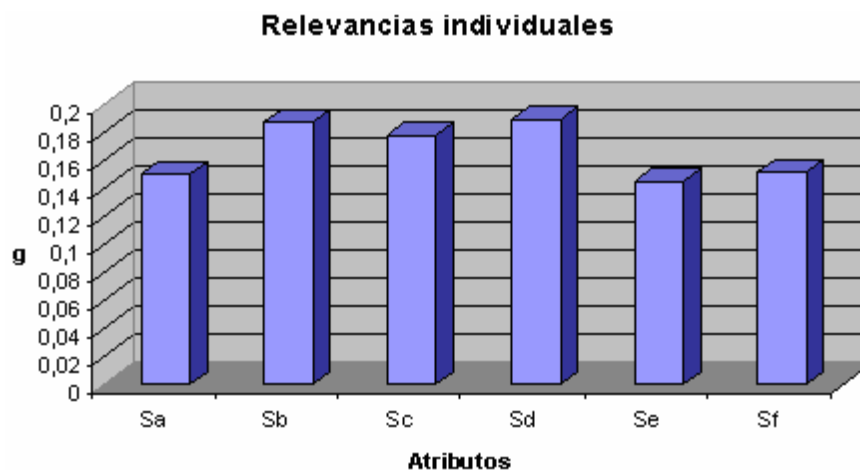


Figura 5.12 Relevancias individuales obtenidas para cada atributo individual a partir del conjunto de pares de imágenes estereoscópicas de entrenamiento.

#### 5.4.1.2 Resultados individuales por atributo y combinados mediante el método CFI

Tal y como se ha descrito en la sección 4.2.4.1, el método *CFI* combina los seis atributos individuales anteriores para tomar una decisión conjunta sobre la elección de un candidato de entre los posibles.

En la figura 5.13 se muestra el mapa de disparidad obtenido para el par estereoscópico dado en las figuras 1.3(a) y (b) mediante la aplicación del método *CFI*, concretamente, en (a) sin utilizar el proceso de segmentación diseñado según se explica en el capítulo tres y en (b) el resultado aplicando dicho proceso. La barra de color representa la asociación existente entre cada color de la imagen y el valor de la disparidad asignada. Obsérvese la variación del mapa en el rango  $[0, 8]$ , tal y como se

determinó con anterioridad. Resulta fácil deducir, a la vista de ambas figuras, que cuando se aplica el proceso de segmentación aparece una mayor claridad, al menos en la parte de texturas excluidas durante la segmentación. Esto supone una mejora cualitativa y cuantitativa importante, ya que en un futuro y como continuación de este trabajo de investigación, cuando haya que continuar el proceso para la identificación de los troncos de los árboles, esto será mucho más fácil a partir de las imágenes del tipo (b) que a partir de aquellas otras como en (a). Además de esta mejora cualitativa es fácil deducir la existencia de una mejora cuantitativa por el mero hecho de que algunos píxeles pertenecientes a texturas no excluidas por la ausencia de segmentación, se emparejan erróneamente con píxeles pertenecientes a los troncos, cosa que no ocurre cuando dichas texturas se excluyen, ya que no se les ofrece esa posibilidad. El porcentaje de error se incrementa aproximadamente un 9-10 % cuando no se aplica segmentación.

Si en lugar de combinar los atributos se utiliza cada uno de ellos individualmente, de forma que entre dos píxeles procedentes de las dos imágenes del par estereoscópico se mide la similitud en base a un atributo individual, siempre es posible elegir de entre varios candidatos por similitud. Aplicando por tanto, el protocolo de correspondencia descrito en la misma sección 4.2.4.1 para cada uno de los atributos individuales, podemos obtener el correspondiente mapa de disparidad y de ahí computar los errores cometidos por comparación con el *ground truth* generado para cada par estereoscópico del conjunto disponible.

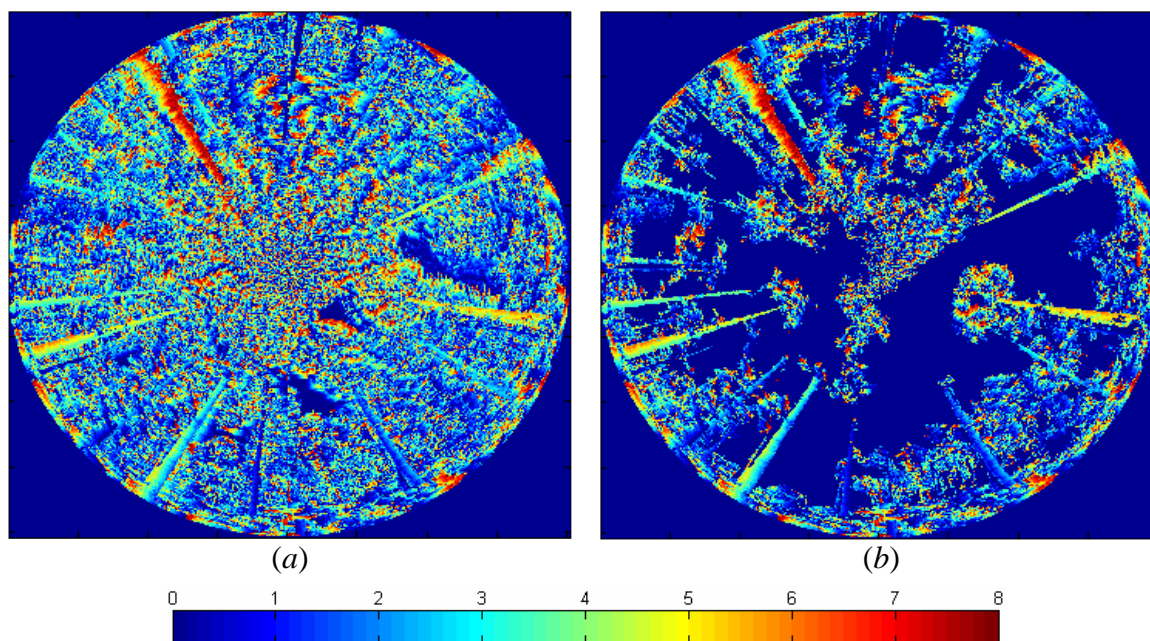


Figura 5.13 Mapa de disparidad inicial obtenido con la propuesta *CFI*: (a) sin segmentación; (b) con segmentación según el proceso combinado



Debido a que el enfoque *CFI* es en realidad una estrategia combinada basada en la teoría de la toma de decisiones, se ha elegido para su comparación el método clásico de Yager (1988) basado en la combinación de medidas *fuzzy*. En el caso que nos ocupa se obtiene una medida de similitud final combinando dos a dos las medidas de similitud individuales, según el siguiente operador:

$$S_i(l) = 1 - \min \left\{ 1, \left( (1 - s_{ih})^p + (1 - s_{ik})^p \right)^{\frac{1}{p}} \right\}, \quad p \geq 1 \quad (5.1)$$

donde  $h$  y  $k$  denotan las dos medidas de similitud involucradas. A continuación, aplicando la propiedad asociativa de este operador *suma* se calcula el grado de similitud proporcionado por las seis medidas de similitud individuales. El parámetro  $p$  se estima a partir del mismo conjunto de imágenes estereoscópicas utilizado para el entrenamiento, que a su vez han permitido obtener la relevancia de los atributos individuales. El procedimiento para fijar el parámetro  $p$  consiste en tomar como referencia los correspondientes *ground truth* para cada una de las imágenes destinadas a este propósito, de forma que variando  $p$  de 1 a 4 se computa el porcentaje de error para cada uno de esos valores, eligiendo finalmente el valor de  $p$  con el que se obtiene el mínimo error, que resulta ser en nuestro caso igual a 2.0.

En definitiva, tomando el conjunto de pares estereoscópicos disponibles para las pruebas y aplicando para cada par los seis criterios individuales por separado se obtienen seis mapas de disparidad, para dicho par, esto es, uno por criterio. Mediante el operador de Yager (1988) se obtiene también un mapa para cada par de imágenes estereoscópicas y finalmente, tras aplicación del método *CFI* se obtiene igualmente un nuevo mapa de disparidad por cada par estereoscópico. Con cada mapa de disparidad y tomando como referencia el correspondiente *ground truth* para el par bajo análisis, se computa el error obtenido para ese par de imágenes estereoscópicas. Finalmente, promediando los errores para el conjunto de dieciséis pares disponibles se obtienen los resultados mostrados en la tabla 5.6, donde se muestra el valor promedio del error junto con sus desviaciones estándar medias según la similitud proporcionada por los seis atributos individuales, identificados bajos las columnas:  $(s_a, s_b, s_c, s_d, s_e, s_f)$ . En dicha tabla se muestran también los porcentajes promediados de error y sus desviaciones estándar promediadas sobre el mismo conjunto de los dieciséis pares de imágenes estereoscópicas según el método de Yager (*YAG*) y la propuesta combinada *CFI*.

Tabla 5.6 Porcentajes promediados de error y desviaciones estándar según los seis atributos individuales, el método combinado de Yager y la propuesta basada en la toma de decisiones *CFI*.

Porcentajes promediados de error y desviaciones estándar															
$s_a$		$s_b$		$s_c$		$s_d$		$s_e$		$s_f$		YAG		CFI	
%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$
30.1	2.9	16.2	1.3	18.1	1.7	14.3	1.1	35.2	3.6	32.1	3.1	13.3	1.9	<b>11.2</b>	<b>1.3</b>

A la vista de los resultados mostrados en la tabla 5.6 se deduce fácilmente que para el conjunto de imágenes analizadas, las mejores estrategias son las que utilizan los atributos combinados frente a las que los aplican por separado. Algo que por otra parte no supone ninguna sorpresa, como viene siendo puesto de manifiesto a través de diversos estudios de investigación, principalmente en el área del reconocimiento de patrones (Kuncheva, 2004). Dentro de las dos propuestas combinadas, resulta ser la *CFI* la que obtiene los mejores resultados. Se observa igualmente que las desviaciones estándar también son menores para la estrategia *CFI*, lo que significa que existe una menor dispersión de los errores respecto de la media cuando se utiliza esta técnica. Puesto que tanto *YAG* como *CFI* combinan los mismos seis atributos, la razón de esta mejora estriba necesariamente en el hecho de que *CFI* conlleva un proceso orientado a la obtención de las relevancias individuales de cada atributo, proceso éste que puede encuadrarse dentro de las técnicas de aprendizaje con las correspondientes ventajas asociadas.

Por otra parte, a partir de los resultados obtenidos, también puede fácilmente deducirse, que los mejores resultados respecto de la aplicación de los criterios individuales se obtienen a través de las similitudes proporcionadas por el atributo módulo del gradiente ( $s_d$ ), coincidiendo con el hecho de que este mismo criterio ha sido el mejor valorado desde el punto de vista de su mayor relevancia, tal y como se ha puesto de manifiesto en la sección 5.4.1.1, encontrándose así una relación directa entre relevancia y resultados, de suerte que a mayor relevancia mayor porcentaje de aciertos.

Para mayor facilidad en la interpretación de los resultados mostrados en la tabla 5.6, en las figuras 5.14 y 5.15 se muestran gráficamente los mismos resultados en porcentajes promediados de error, figura 5.14 y desviaciones estándar, figura 5.15, remarcándose en rojo el menor porcentaje de error y la menor desviación estándar obtenidas mediante el método *CFI*.

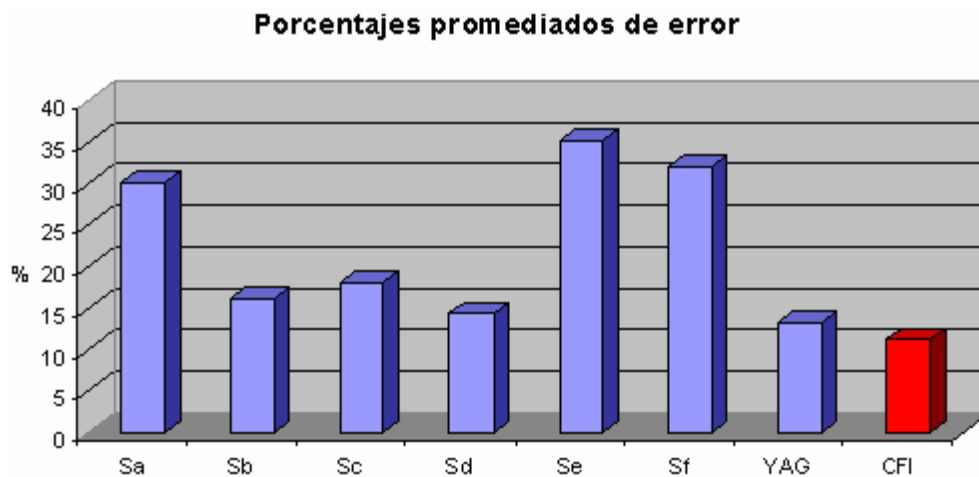


Figura 5.14 Porcentajes promediados de error obtenidos por aplicación de los criterios por separado y en combinación a través de los métodos *YAG* y *CFI*.



Figura 5.15 Desviaciones estándar promediadas obtenidas por aplicación de los criterios por separado y en combinación a través de los métodos *YAG* y *CFI*.

### 5.4.2 Integral Fuzzy de Sugeno (SFI)

Según se recoge en la sección 4.2.4.2, el método *SFI* es en esencia similar al *CFI* en cuanto a su estrategia, involucrando los mismos seis atributos y las densidades *fuzzy*, que se obtienen a partir de las relevancias de cada criterio. Dichas relevancias son, por tanto, las mismas que las mostradas en la tabla 5.5. La única diferencia estriba en la forma de obtener el grado de correspondencia, que en el caso del *SFI* se lleva a cabo mediante la ecuación (4.9) en lugar de la (4.8), que es la utilizada en el *CFI*. Se trata pues de una mínima diferencia entre ambos métodos, que nos conduce a la obtención de los mismos resultados en ambos casos. Es decir, el porcentaje medio de error obtenido sobre el conjunto de imágenes de test, resulta ser del 11.2% con una desviación estándar de 1.3, al igual que en el *SFI*.

La razón que explica la obtención de similares resultados la podemos encontrar en el uso de las densidades *fuzzy* y no así en el cómputo del grado de correspondencia. Es decir, en la importancia asignada a cada atributo en base a los cuales se toma la decisión final.

Ante esta perspectiva los resultados obtenidos respecto de los mapas de disparidad son prácticamente los mismos, sin apreciarse diferencia respecto de los obtenidos mediante el *CFI* y en particular los mostrados en la figura 5.13 para el mismo par de imágenes estereoscópicas.

### 5.4.3 Teoría de Dempster-Shafer (DES)

#### 5.4.3.1 Cálculo de los vectores de medias

Tal y como se expuso en la sección 4.2.4.3, el método *DES* propuesto en este trabajo vuelve a ser en realidad una adaptación de un método combinado de clasificadores, que utiliza plantillas para la asignación de píxeles bien a la clase de correspondencias correctas o incorrectas, que previamente se han establecido a partir del conjunto de cuatro pares de imágenes estereoscópicas destinadas a tal efecto. Este conjunto es el mismo que el utilizado previamente en *CFI* o *SFI* para estimar las relevancias relativas de los criterios. Por tanto, en realidad se dispone de pares de píxeles clasificados como pertenecientes a una de las dos clases creadas al efecto:  $w_1$ , correspondencias positivas o ciertas y  $w_2$ , correspondencias negativas o falsas.

El número de muestras pertenecientes a la clase  $w_1$  es de 2000 pares de píxeles seleccionados aleatoriamente de entre los obtenidos como correspondencias verdaderas, tras la comprobación de esta hipótesis manualmente. Por otra parte, el número de muestras pertenecientes a la clase  $w_2$  es también de 2000, igualmente seleccionadas aleatoriamente del conjunto de pares de píxeles que han resultado falsas correspondencias tras su verificación manual.

Como se mencionó en la misma sección 4.2.4.3, las muestras están representadas por vectores de dimensión seis, donde las componentes son los valores de similitud correspondientes a cada uno de los seis criterios disponibles, provenientes de otros tantos atributos y consecuentemente en el mismo orden previamente establecido, es decir:  $(s_a, s_b, s_c, s_d, s_e, s_f)$ .

A partir de las muestras asignadas a cada clase se calculan los vectores representativos de las mismas, resultando ser los vectores de medias asociados con cada una de las clases. La tabla 5.7 muestra los valores de las medias  $\bar{\mathbf{v}}_j = [\bar{s}_{ja}, \bar{s}_{jb}, \bar{s}_{jc}, \bar{s}_{jd}, \bar{s}_{je}, \bar{s}_{jf}]^T$  obtenidas con las muestras anteriores para cada una de las dos clases.

Tabla 5.7 Vectores de medias  $\bar{\mathbf{v}}_j$  obtenidos para las clases de correspondencias positivas ( $w_1$ ) y negativas ( $w_2$ ) a partir del conjunto de pares de imágenes de entrenamiento.

$\bar{\mathbf{v}}_j$	
$\bar{\mathbf{v}}_1$ (clase $w_1$ )	$\bar{\mathbf{v}}_2$ (clase $w_2$ )
$[0.81, 0.85, 0.92, 0.96, 0.79, 0.80]^T$	$[0.12, 0.11, 0.08, 0.07, 0.11, 0.10]^T$

Las componentes de los vectores representativos de las muestras se encuentran normalizadas en el rango  $[0,1]$ , destacando el hecho de que a mayor similitud entre los pares de píxeles mayor es su proximidad al valor unidad y viceversa. En la figura 5.16 se representan gráficamente los dos vectores de medias conjuntamente, poniéndose de manifiesto el hecho de que las componentes del vector  $\bar{\mathbf{v}}_1$  tienden hacia uno, mientras que las del vector  $\bar{\mathbf{v}}_2$  tienen una clara tendencia hacia cero. Esto significa que se han elegido pares de correspondencias claramente positivas y negativas, con el fin de evitar el mayor número de ambigüedades posible durante la fase de decisión del posterior proceso de correspondencia, que se describe a continuación.

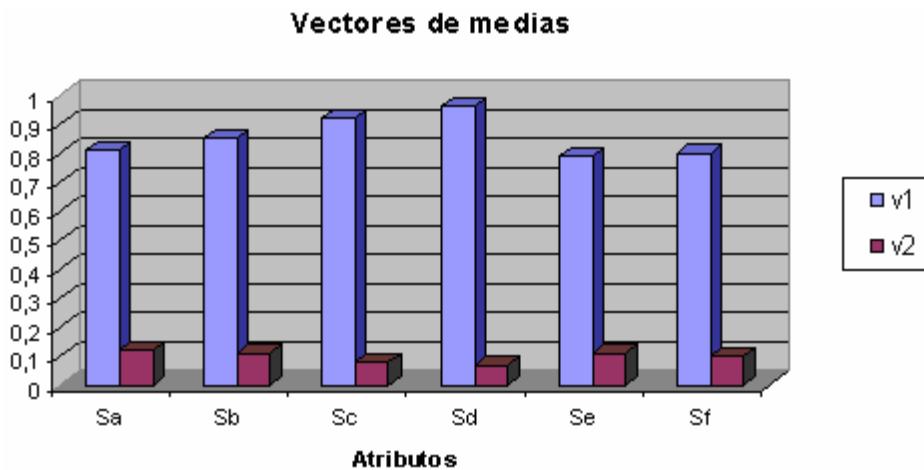


Figura 5.16 Vectores de medias obtenidos para las clases de correspondencias positivas y negativas a partir del conjunto de pares de imágenes de entrenamiento.

### 5.4.3.2 Resultados obtenidos mediante el método DES

Siguiendo el mismo protocolo que el descrito para los métodos *CFI* y *SFI*, es decir tratando de emparejar píxeles de izquierda a derecha y de derecha a izquierda según las imágenes de referencia, se aplica la ecuación (4.12) para establecer los grados de correspondencia entre los píxeles bajo emparejamiento. De nuevo, tal y como puede deducirse fácilmente a partir de dicha ecuación, intervienen los mismos seis atributos mediante las similitudes correspondientes, si bien en este caso las correspondencias se establecen por comparación con los vectores representativos de cada clase, esto es  $\bar{v}_1$  y  $\bar{v}_2$  estimados previamente.

Utilizando el mismo conjunto de dieciséis pares de imágenes de prueba que para los métodos *CFI* y *SFI* se obtiene el mismo porcentaje medio de error, es decir, el 11.2%. No obstante, se observa un ligero incremento en el valor de la desviación estándar, que pasa a ser de 1.6. Esto se debe probablemente a las muestras representativas de cada una de las clases, ya que diversos experimentos llevados a cabo variando el número de muestras nos confirman siempre el mismo comportamiento. No obstante no se puede deducir con claridad el origen de esa ligera desviación, cabiendo incluso la posibilidad de que sea el propio método el origen de la misma, que por otra parte no resulta ser especialmente significativa. En cualquier caso, el mapa de disparidad así obtenido no presenta diferencias observables con respecto al mostrado en la figura 5.13.

## 5.4.4 Toma de Decisiones Multicriterio Fuzzy (MCDMF)

### 5.4.4.1 Cálculo de los pesos asociados a cada criterio

En este método, tal y como se ha descrito en la sección 4.2.4.4, se dispone de dos criterios y a cada criterio se le asigna un peso específico determinado, estimándose a partir del conjunto de cuatro pares de imágenes estereoscópicas disponibles a tal efecto. Estas imágenes son las mismas que las utilizadas en los tres métodos precedentes, de forma que en el caso que nos ocupa, dichos pesos se calculan de la misma forma que las relevancias relativas en los métodos *CFI* y *SFI*, obteniendo por tanto los mismos valores que para éstas. En la tabla 5.8 se reproducen los valores de la tabla 5.5, expresando explícitamente los valores de los pesos específicos asignados a

cada atributo. Por este motivo, la figura 5.12 es también una representación gráfica de los pesos específicos mostrados en la tabla 5.8.

Tabla 5.8 Porcentajes promediados de error y pesos específicos obtenidos para cada atributo con el conjunto de pares de imágenes de entrenamiento.

Porcentajes promediados de error y pesos específicos de los criterios		
Criterio	$p_h$	$w_h$
$s_a$	28	0.150
$s_b$	10	0.187
$s_c$	14	0.178
$s_d$	9	0.189
$s_e$	30	0.145
$s_f$	27	0.151

A partir de los valores anteriores se construyen los vectores de pesos involucrados en el método, a saber:  $w_1 = (w_a, w_b, w_c) = (0.150, 0.187, 0.178)$  y  $w_2 = (0.189, 0.145, 0.151)$ . En realidad, los pesos específicos son números triangulares *fuzzy*, por lo que cada uno de estos números está formado por una terna, tal y como se definió en el capítulo cuatro. Como no podría ser de otra manera, el mayor peso específico coincide con la mayor relevancia relativa, correspondiéndose en ambos casos con el atributo módulo del gradiente.

#### 5.4.4.2 Resultados obtenidos mediante el método MCDMF

Siguiendo de nuevo el mismo protocolo que el descrito para los tres métodos anteriores, *CFI*, *SFI* y *DES*, de nuevo ahora se trata de emparejar bien los píxeles de la imagen izquierda con los de la derecha, o los de la derecha con los de la izquierda, en este caso bajo el paradigma del *MCDMF* descrito en la sección 4.2.4.4, y de acuerdo con el coeficiente de proximidad definido en la ecuación (4.18). De nuevo, tal y como puede deducirse fácilmente a partir de dicha ecuación, intervienen los mismos seis atributos a través de sus similitudes individuales, donde cada atributo tiene asociado un peso específico, cuyos valores son los obtenidos en la tabla 5.8.

Utilizando el mismo conjunto de imágenes de test que para los tres métodos previos, *CFI*, *SFI* y *DES*, se obtiene un porcentaje medio de error que resulta ser en este caso del 9.3% con una desviación estándar de 0.9.

En la figura 5.17 se muestra el mapa de disparidad obtenido mediante la aplicación del procedimiento *MCDMF*. Como fácilmente puede observarse, las diferencias entre este mapa y el que aparece en la figura 5.13(b) obtenido, en este caso, mediante el método *CFI*, son realmente mínimas, resultando muy difícil su distinción visual. Se observa principalmente que a algunos píxeles procedentes de las texturas que eran objeto de exclusión durante la segmentación, se les ha asignado valores de disparidad nulos, lo que facilitará en un futuro el análisis de las texturas de los troncos.

Para facilitar el análisis de resultados en conjunto, en la tabla 5.9 se recopilan los porcentajes medios y las desviaciones estándar obtenidos con cada uno de los cuatro métodos propuestos en este trabajo. No se incluyen los resultados de los métodos individuales ni los obtenidos por *YAG* por haber sido superados claramente por el resto de métodos. Los mismos resultados mostrados en la tabla 5.9 se representan gráficamente en las figuras 5.18(a) y (b) donde aparecen los porcentajes de error promediados y las desviaciones estándar respectivamente.

Tabla 5.9 Porcentajes promediados de error y desviaciones estándar para los métodos: *CFI*, *SFI*, *DES* y *MCDMF*.

Porcentajes promediados de error y desviaciones estándar							
<i>CFI</i>		<i>SFI</i>		<i>DES</i>		<i>MCDMF</i>	
%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$
<b>11.2</b>	<b>1.3</b>	<b>11.2</b>	<b>1.3</b>	<b>11.2</b>	<b>1.6</b>	<b>9.3</b>	<b>0.9</b>

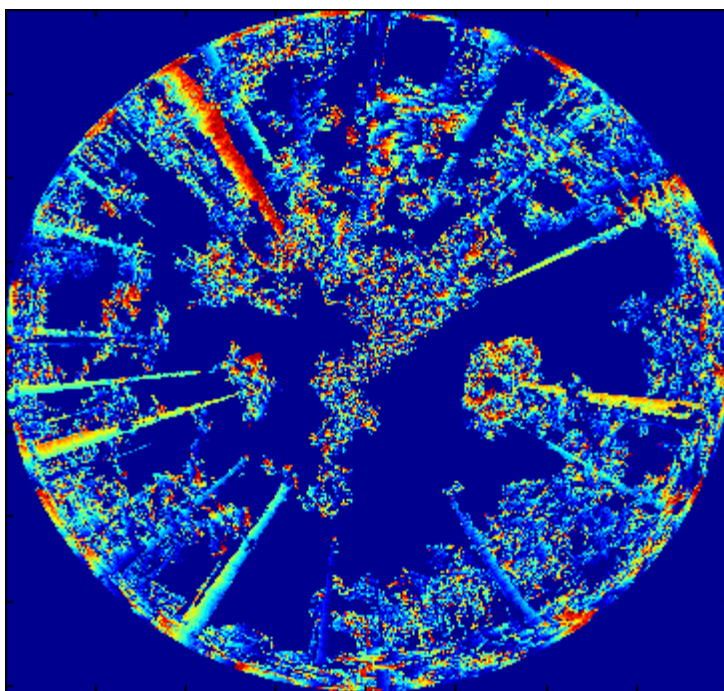


Figura 5.17 Mapa de disparidad inicial obtenido con la propuesta *MCDMF*.



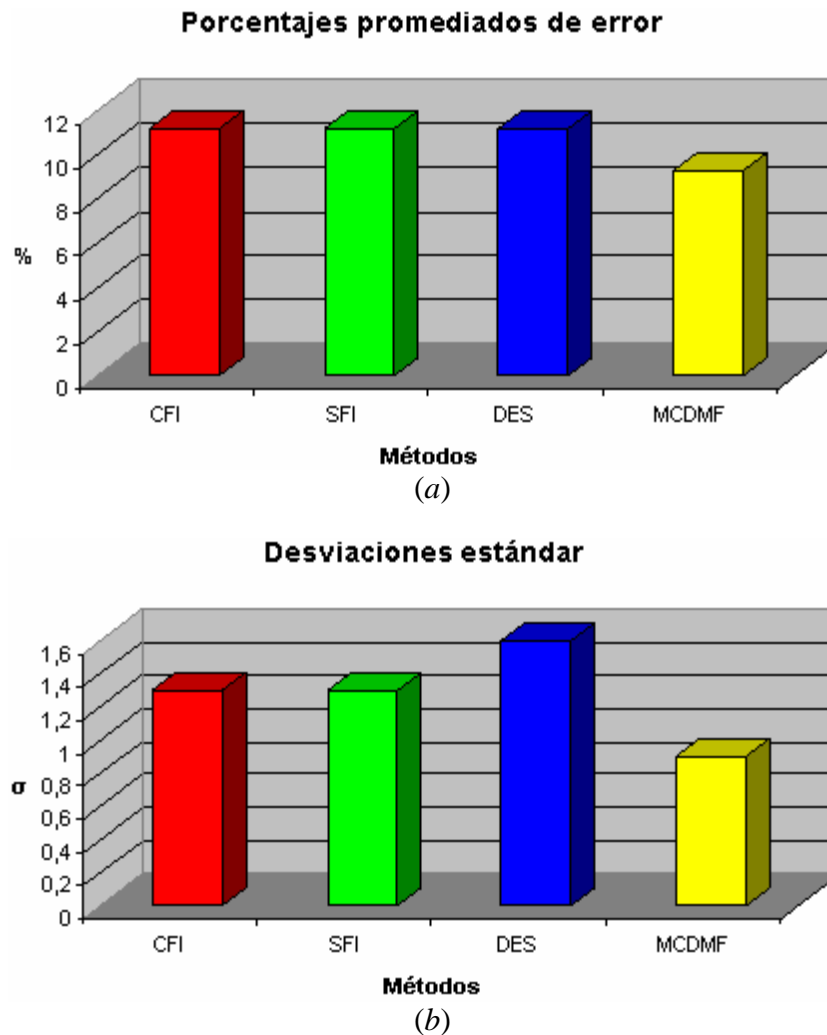


Figura 5.18 (a) Porcentajes promediados y (b) desviaciones estándar obtenidas por los métodos *CFI*, *SFI*, *DES* y *MCDMF*.

A la vista de los resultados mostrados en la tabla 5.9 y las figuras asociadas 5.18(a) y (b), resulta claro el mejor comportamiento obtenido por el método *MCDMF*, esto es así a pesar de que en todos ellos se están utilizando los mismos criterios de similitud para tomar las decisiones y que en los cuatro casos se utilizan los métodos previstos para el cómputo bien de las relevancias relativas de los atributos (*CFI*, *SFI*), bien de los vectores representativos de las clases de correspondencias verdaderas y falsas (*DES*), o bien los pesos específicos asociados con los criterios y por tanto con los atributos (*MCDMF*). Es más, los pesos específicos y las relevancias resultan ser idénticos, tal y como se refleja en las tablas 5.5 y 5.8. Esto nos lleva a pensar que la mejora de resultados obtenida con el método *MCDMF* proviene de la mayor flexibilidad que proporciona dicho método frente a *CFI* y *SFI*, dada la codificación de los datos como números triangulares *fuzzy*. Diversos experimentos llevados a cabo en el sentido de intercambiar los atributos en los números triangulares *fuzzy* definidos a tal

efecto,  $x_{i1} = (s_{ia}, s_{ib}, s_{ic})$  y  $x_{i2} = (s_{id}, s_{ie}, s_{if})$  no han aportado ninguna evidencia en relación a que sea éste el origen de la mejora, ya que en la mayoría de los casos los resultados no mejoran sino que llegan incluso a empeorar.

Una observación digna de mención que conviene añadir a este análisis es el hecho de que cuando no se aplica el proceso de segmentación descrito en el capítulo tres (con los resultados mostrados en la sección 5.2 para el caso de las correspondencias basadas en píxeles) los resultados empeoran considerablemente, llegando a incrementarse el promedio del error en un porcentaje similar al del *CFI*, es decir, alrededor del 10%. Esto es debido a que existe un cierto número de píxeles, equivalentes a ese porcentaje, y pertenecientes a los troncos que se emparejan erróneamente con texturas carentes de interés (cielo, hierba del suelo u hojas de los pinos) por el hecho de no haberse excluido mediante el proceso de segmentación.

Como colofón al análisis previo, en lo que sigue siempre se realiza el proceso de segmentación previsto, siendo el mapa de disparidad generado por el método *MCDMF* el que finalmente se utiliza como entrada para su filtrado mediante los procedimientos basados en los Mapas Cognitivos *Fuzzy* y la Red Neuronal de Hopfield, que se abordan en las dos secciones siguientes.

## 5.5 Resultados del filtrado del Mapa de Disparidad

### 5.5.1 Mapas Cognitivos Fuzzy (FCMs)

El procedimiento *FCM* tiene como finalidad realizar un filtrado del mapa de disparidad, al objeto de obtener un mapa con disparidades más suaves en aquellas zonas donde realmente conviene, cual es el caso de los troncos de los árboles. Fuera de los troncos el filtrado también surte sus efectos, como se verá posteriormente, si bien por las razones expuestas en los capítulos anteriores, en éstas carece de interés. Tal y como se ha descrito en la sección 4.2.6.1, se trata de un proceso iterativo que busca la estabilidad de la red mediante la modificación de los estados de los nodos. Se ha verificado sobre el conjunto de imágenes analizadas que en promedio se alcanza la convergencia con tan sólo dos iteraciones, es más, en algunos pares con una única iteración es suficiente. Esto hace que los valores de los estados y en consecuencia los valores de disparidad que representan se modifiquen rápidamente como consecuencia de

alcanzar la convergencia con ese mínimo número de iteraciones. Tomando como entrada el mapa de disparidad proporcionado por el método *MCDMF*, representado por la imagen de la figura 5.17, tras la aplicación del proceso con dos iteraciones se obtiene el mapa de disparidad suavizado de la imagen de la figura 5.19.

Por otro lado, se ha constatado también de forma general sobre el conjunto de imágenes analizadas, que cuando se fuerza el proceso al aumentar el número de iteraciones aunque se haya conseguido la convergencia, se produce un suavizado excesivo con una cierta tendencia a modificar los valores de las disparidades en el sentido de los valores máximos existentes en el entorno de vecindad del píxel sobre el que se está actuando. Esto es debido, fundamentalmente, al hecho de que bajo este procedimiento las disparidades individuales son forzadas en exceso por las influencias externas ejercidas por los vecinos, lo cual ciertamente no constituye un comportamiento deseable.

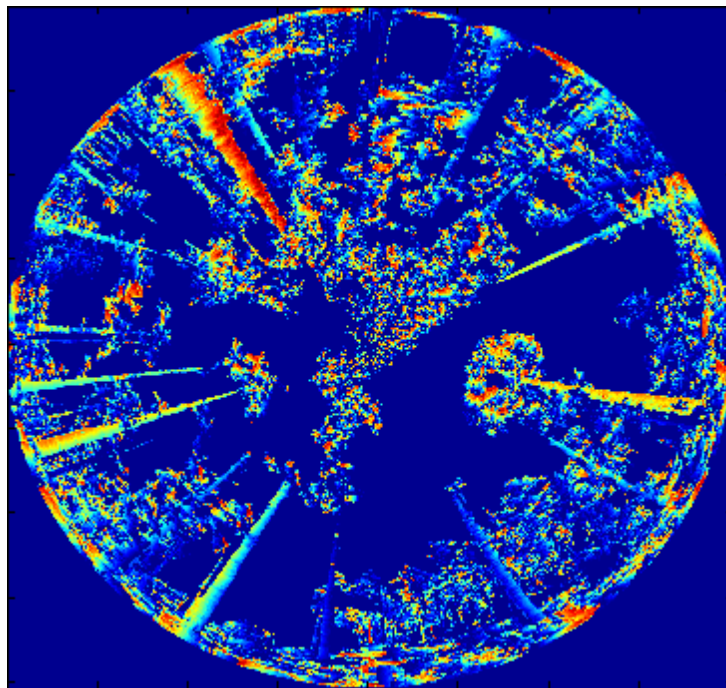
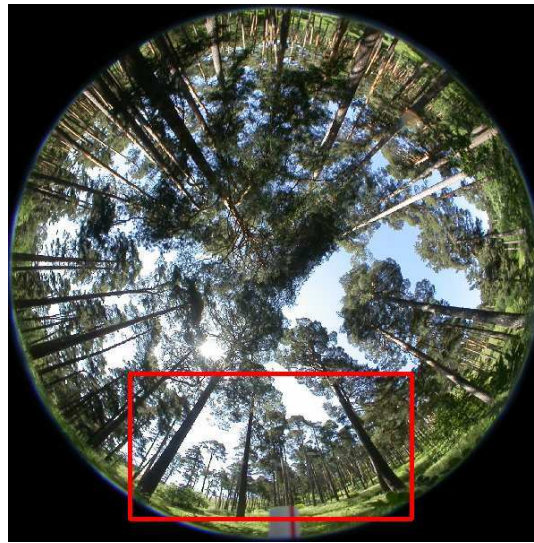


Figura 5.19 Mapa de disparidad obtenido con la propuesta *FCM*.

Para evidenciar más claramente el efecto del suavizado llevado a cabo por el mecanismo *FCM*, delimitamos sobre la imagen de la figura 1.3(a) el área marcada en la figura 5.20(a), que aparece ampliada y expandida en la imagen de la figura 5.20(b). La figura 5.20(c) muestra el mapa de disparidad inicial obtenido mediante el enfoque

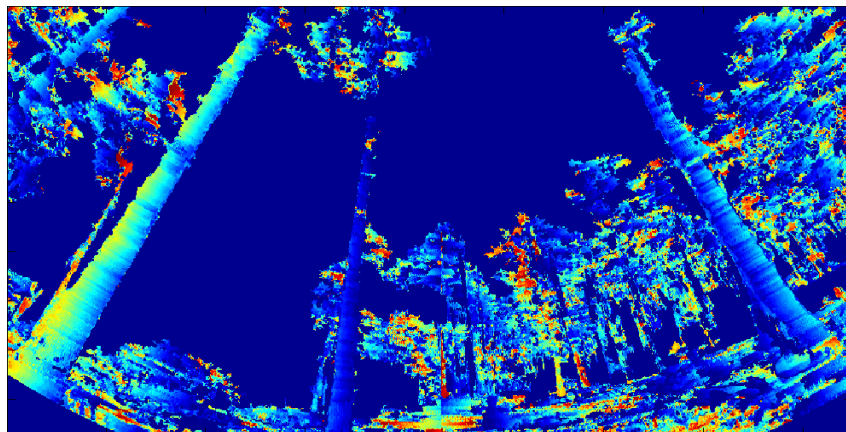
*MCDMF* para dicha área expandida. Finalmente en la figura 5.20(d) se muestra el mapa de disparidad filtrado mediante el enfoque *FCM*.



(a)



(b)



(c)

Figura 5.20 (a) Imagen izquierda omnidireccional; (b) área expandida correspondiente al área marcada en la imagen de la figura 5.20(a); (c) mapa de disparidad obtenido con la propuesta *MCDMF*; (d) mapa de disparidad obtenido con la propuesta *FCM*.

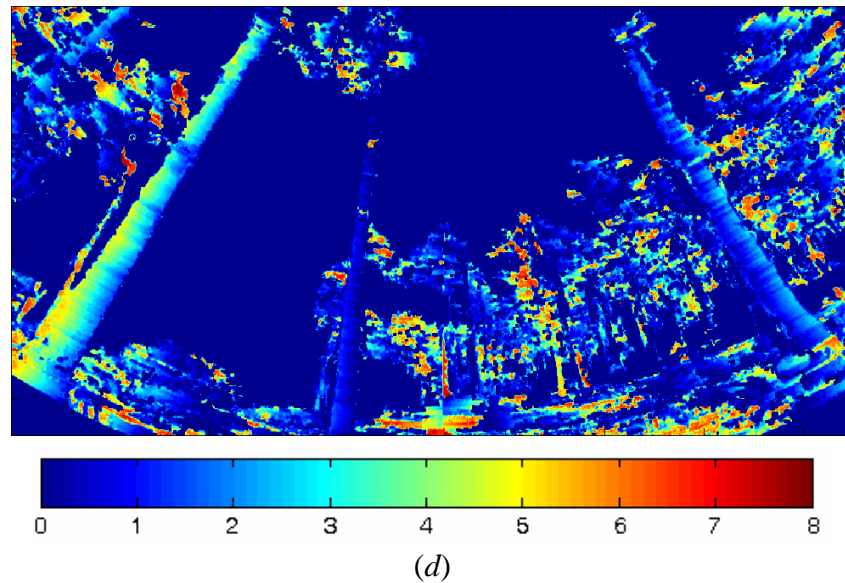


Figura 5.20 Cont.

A la vista de los mapas de disparidad representados en las zonas expandidas, al tronco situado en la parte izquierda del mapa de la figura 5.20(d) le corresponde un nivel promedio de disparidad de 4.69, según la tabla 5.1. Se puede observar en la figura (d) que bastantes niveles de disparidad se mueven ligeramente hacia una tonalidad más amarilla, aproximándose hacia el valor del *ground truth* dado en la mencionada tabla por comparación con la barra de color mostrada en la parte inferior de la figura 5.20(d). Una observación detallada a lo largo de todos los troncos permite llegar a conclusiones similares, si bien sobre distintos niveles de disparidad.

Igualmente, este mismo suavizado se constata sobre los píxeles fuera de los troncos, donde muchas zonas de píxeles aislados se han reducido o incluso desaparecido. Aunque esto en sí mismo no resulta ahora relevante puesto que el interés se centra exclusivamente en los troncos, sí puede resultar útil en el futuro por el hecho de que los propios troncos quedan mejor aislados que antes, de forma que pueden identificarse mejor como regiones de disparidad homogéneas.

Por comodidad, la tabla 5.10 vuelve a reproducir los resultados de porcentajes promediados de error y desviaciones estándar obtenidos por los métodos *CFI*, *SFI*, *DES* y *MCDMF*, para una mejor comparación con los obtenidos mediante el *FCM*. Naturalmente, esos porcentajes de error y las desviaciones estándar se han obtenido promediando los resultados sobre el conjunto de imágenes de test y utilizando los



mismos mapas de referencia (*ground truth*). Para mayor claridad, los resultados de la tabla 5.10 se reproducen gráficamente en las figuras 5.21 y 5.22.

Tabla 5.10 Porcentajes promediados de error y desviaciones estándar obtenidas por el enfoque *FCM* frente a las propuestas de toma de decisiones objeto de investigación.

Porcentajes promediados de error y desviaciones estándar									
<i>CFI</i>		<i>SFI</i>		<i>DES</i>		<i>MCDMF</i>		<i>FCM</i>	
%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$
11.2	1.3	11.2	1.3	11.2	1.6	9.3	0.9	<b>8.6</b>	<b>1.2</b>

Porcentajes promediados de error

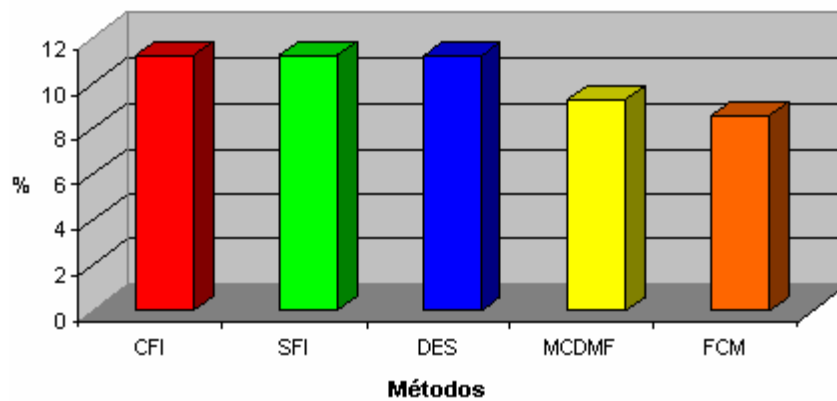


Figura 5.21 Porcentajes promediados de error obtenidos por el enfoque *FCM* frente a los métodos *CFI*, *SFI*, *DES* y *MCDMF*.

Desviaciones estándar

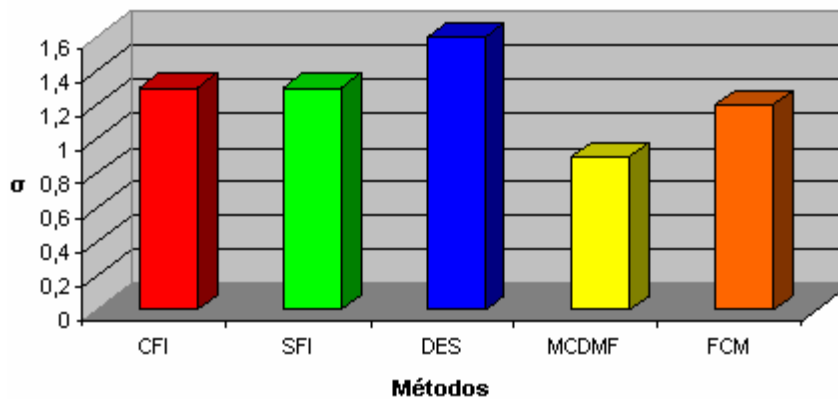


Figura 5.22 Desviaciones estándar obtenidas por el enfoque *FCM* frente a los métodos *CFI*, *SFI*, *DES* y *MCDMF*.

Además de la evidencia cualitativa remarcada previamente sobre los mejores resultados obtenidos por el método *FCM* frente a la mejor de las estrategias previas (*MCDMF*), queda claro también su mejor comportamiento cuantitativo por la disminución del número de errores. No obstante, en relación a la desviación estándar se

ha producido ahora un incremento de dicho valor sobre lo que hasta el momento no se ha podido extraer ninguna consecuencia evidente.

### 5.5.2 Red Neuronal de Hopfield (HNN)

Uno de los motivos que nos condujeron a utilizar la red neuronal de Hopfield fue el hecho de que se trata de un proceso de optimización basado en la minimización de una función de energía. Gracias a ello se puede controlar en cada iteración del proceso el valor de dicha función, lo que permite verificar la evolución de la propia red. Este hecho en sí mismo representa una ventaja con respecto al proceso basado en los *FCMs*, razón por la cual se introdujo en su momento. El objetivo final consiste, como en el caso de los *FCMs*, en realizar un filtrado del mapa de disparidad ya obtenido, lo que realmente se traduce en un suavizado de dicho mapa. Con el fin de comparar los resultados con el *FCM*, se parte del mapa de disparidad proporcionado por el enfoque *MCDMF*, para lo cual se aplica el procedimiento descrito en la sección 4.2.6.2.

El mapa de disparidad obtenido como resultado de dicho proceso, a partir de las imágenes de la figura 1.3, es el que se muestra en la figura 5.23.

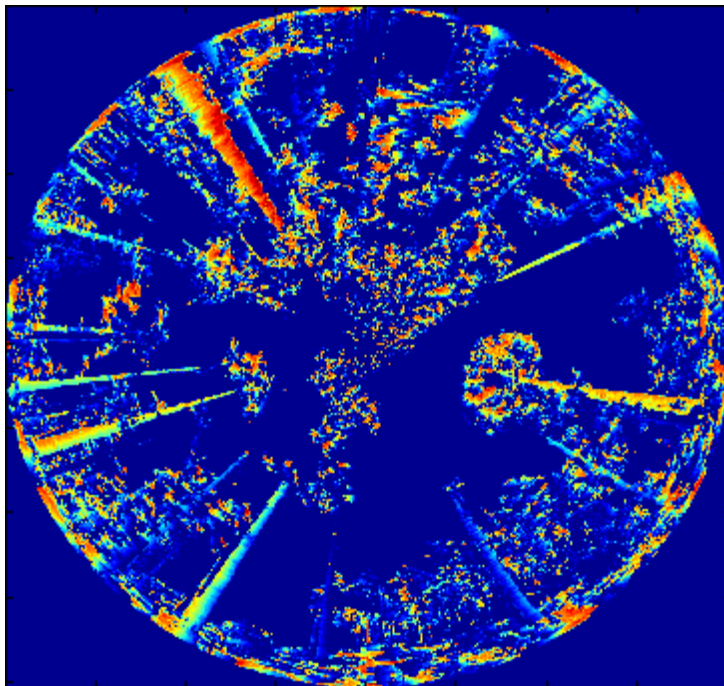
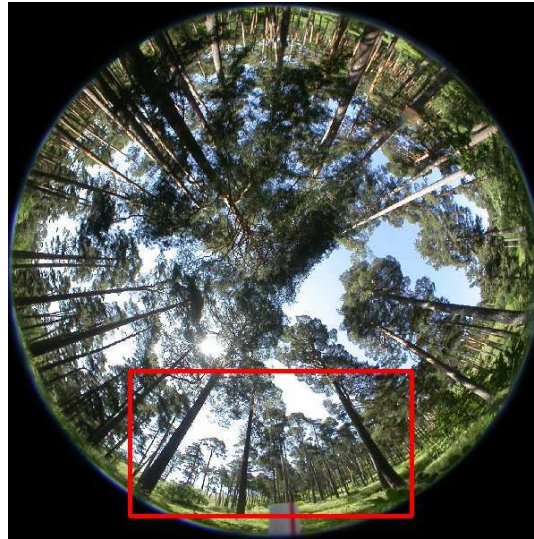


Figura 5.23 Mapa de disparidad obtenido con la propuesta *HNN*.

Por claridad, en la figura 5.24(a) se reproduce la propia figura 1.3(a), señalando sobre ella el área expandida que se muestra en la figura 5.24(b). La figura 5.24(c)

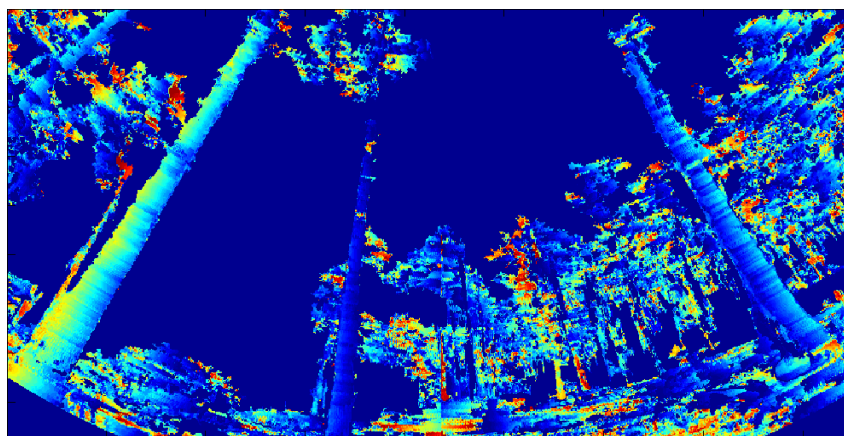
muestra el mapa de disparidad inicial obtenido mediante el enfoque *MCDMF* para el área expandida de la figura 5.24(b). Finalmente en la figura 5.24(d) aparece el mapa de disparidad suavizado mediante el enfoque *HNN*.



(a)



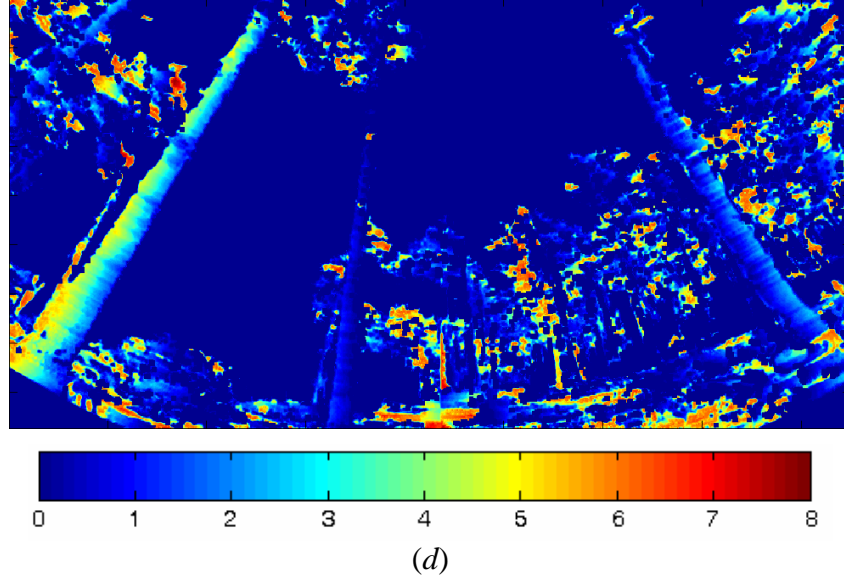
(b)



(c)

Figura 5.24 (a) Imagen izquierda omnidireccional; (b) área expandida correspondiente al área señalada en la imagen de la figura 5.24(a); (c) mapa de disparidad obtenido mediante la propuesta *MCDMF*. (d) mapa de disparidad obtenido con la propuesta *HNN*.



Figura 5.24 *Cont.*

Las mismas conclusiones, a nivel cualitativo expuestas en relación al comportamiento del método *FCM* frente al *MCDMF* son ahora aplicables al caso de los resultados obtenidos por la estrategia *HNN* en relación al mapa de disparidad. Así los niveles de disparidad del tronco que aparece en la parte izquierda de la imagen de la figura 5.24(d) se aproximan como antes al valor medio proporcionado por el *ground truth* y su extensión en relación al resto de troncos de la imagen, que es a su vez extensible al resto de las imágenes de prueba.

También conviene observar el hecho de que en el mapa de la figura 5.24(d), valores de disparidad aislados tanto dentro como fuera de los troncos han cambiado hacia los valores de sus vecinos en relación al mapa de la figura 5.24(c).

Por otra parte, durante los experimentos llevados a cabo con el conjunto de imágenes de test disponibles, se ha observado un comportamiento con tendencias paralelas entre la evolución de la función de energía, figura 5.25(a), y los porcentajes de error, figura 5.25(b), ambos promediados sobre el conjunto de dieciséis pares de imágenes de prueba disponibles y en relación con el número de iteraciones relativas al proceso *HNN*. En dichas figuras la iteración cero se corresponde con el inicio del proceso y por tanto con los datos obtenidos a partir del mapa de disparidad proporcionado por el procedimiento *MCDMF*.

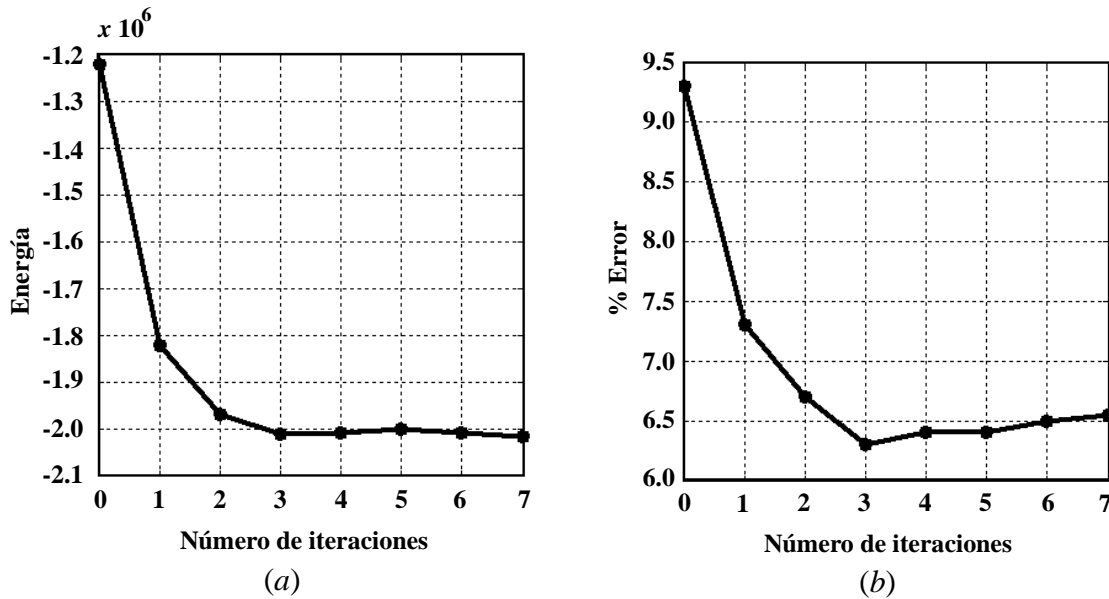


Figura 5.25 (a) Variación promedio de la función de energía con el número de iteraciones; (b) variación del porcentaje promediado de error frente al número de iteraciones.

Obsérvese, en ambos casos, cómo a partir del inicio, correspondiéndose con la iteración cero, se produce un fuerte descenso hasta la primera iteración; a partir de la cual las pendientes se suavizan hasta la iteración tres en la que se obtiene el mínimo valor en el porcentaje de error. La energía, aunque no consigue el mínimo valor hasta la iteración siete, a partir de la tres sigue una trayectoria más o menos estable. Por tanto, a efectos de comparación hemos considerado, como porcentaje promediado del error para *HNN*, el mínimo valor, resultando ser del 6.3%. A partir de la cuarta iteración se produce un sobre-suavizado del mapa de disparidad, de forma que los valores de la disparidad tienden ligeramente a incrementarse, tal y como ocurría en el *FCM* en la segunda iteración.

La tabla 5.11 muestra los porcentajes promediados de error y desviaciones estándar promediadas, que se han obtenido mediante el método *HNN* frente a los obtenidos con las propuestas de toma de decisiones objeto de investigación: *CFI*, *SFI*, *DES* y *MCDMF*. Las gráficas de las figuras 5.26 y 5.27 muestran, respectivamente, los porcentajes promediados de errores y desviaciones estándar.

Tabla 5.11 Porcentajes promediados de error y desviaciones estándar obtenidas por el enfoque *HNN* frente a las propuestas de toma de decisiones objeto de investigación.

Porcentajes promediados de error y desviaciones estándar medias									
<i>CFI</i>		<i>SFI</i>		<i>DES</i>		<i>MCDMF</i>		<i>HNN</i>	
%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$	%	$\sigma$
11.2	1.3	11.2	1.3	11.2	1.6	9.3	0.9	<b>6.3</b>	<b>0.8</b>

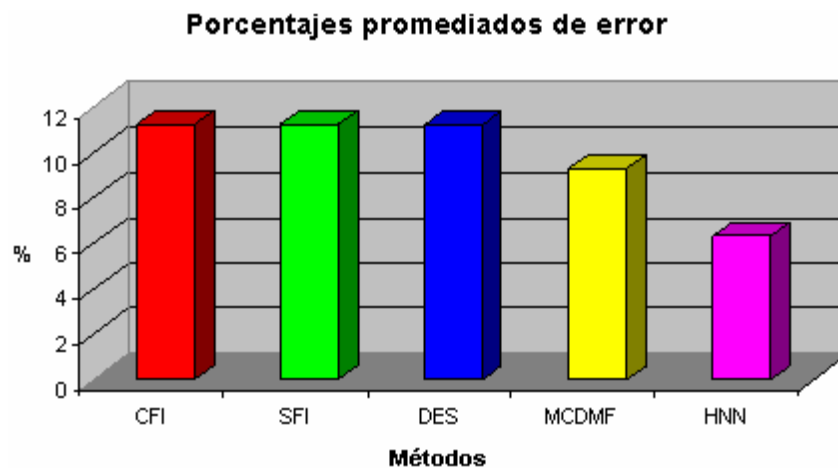


Figura 5.26 Porcentajes promediados de error obtenidos por el enfoque *HNN* frente a cada una de las propuestas de toma de decisiones objeto de investigación: *CFI*, *SFI*, *DES* y *MCDMF*.

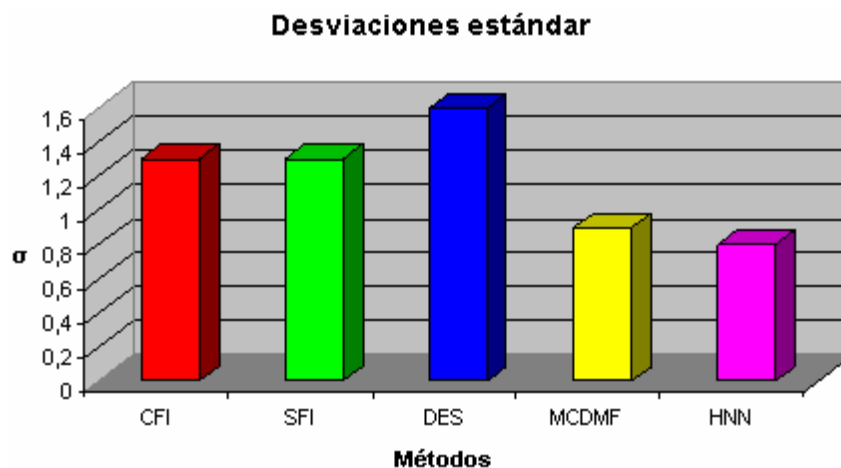


Figura 5.27 Desviaciones estándar obtenidas por el enfoque *HNN* frente a cada una de las propuestas de toma de decisiones objeto de investigación: *CFI*, *SFI*, *DES* y *MCDMF*.

A diferencia de lo que ocurre en el caso del *FCM*, donde la desviación estándar sufre un ligero aumento, en el caso del *HNN* desciende con respecto al resto de métodos. Tampoco resulta fácil concretar el origen de este descenso como ocurría en el caso del *FCM* con respecto al aumento que allí se producía. Esto, añadido al hecho de que también el porcentaje medio de error disminuye, nos permite concluir que la mejor estrategia resulta ser la combinación del método *MCDMF* con el filtrado del mapa de disparidad mediante el *HNN*.

## 5.6 Métodos basados en píxeles: resumen de resultados

A modo de resumen, en la tabla 5.12 se realiza un compendio general de resultados de todos los métodos analizados hasta ahora, y orientados al establecimiento de las correspondencias mediante píxeles. En primer lugar aparecen los métodos que

utilizan los atributos de los píxeles de forma individual, identificándose de la forma habitual como  $s_a$ ,  $s_b$ ,  $s_c$ ,  $s_d$ ,  $s_e$ ,  $s_f$ , según el atributo que representan. En segundo lugar, se muestran los resultados de los métodos que combinan dichos atributos mediante diferentes estrategias, entre ellas se incluye la clásica de Yager (1988), *YAG*, así como las cuatro propuestas en este trabajo, identificadas como *CFI*, *SFI*, *DES* y *MCDMF*. Finalmente se muestran los resultados obtenidos tras el filtrado del mapa de disparidad mediante los métodos *FCM* y *HNN*, que tienen como punto de partida el mapa de disparidad proporcionado por *MCDMF*, que es el que ha obtenido los mejores resultados en relación tanto con los métodos individuales como con los combinados.

Tabla 5.12 Porcentajes promediados de error y desviaciones estándar obtenidas para cada atributo por separado y para los métodos *YAG*, *CFI*, *SFI*, *DES*, *MCDMF*, *FCM* y *HNN*, resaltando en negrita los mejores resultados en cada categoría.

Porcentajes promediados de error y desviaciones estándar <i>medias</i>			
<b>Categoría</b>	<b>Criterios/ métodos</b>	<b>%</b>	<b><math>\sigma</math></b>
<b>Atributos individuales</b>	$s_a$ (correlación)	30.1	2.9
	$s_b$ (color)	16.2	1.3
	$s_c$ (textura)	18.1	1.7
	$s_d$ (módulo gradiente)	<b>14.3</b>	<b>1.1</b>
	$s_e$ (ángulo gradiente)	35.2	3.6
	$s_f$ (Laplaciana)	32.1	3.1
<b>Estrategias combinadas basadas en similitud</b>	<i>YAG</i>	13.3	1.9
	<i>CFI</i>	11.2	1.3
	<i>SFI</i>	11.2	1.3
	<i>DES</i>	11.2	1.6
	<i>MCDMF</i>	<b>9.3</b>	<b>0.9</b>
<b>Filtrado del mapa de disparidad</b>	<i>FCM</i>	8.6	1.2
	<i>HNN</i>	<b>6.3</b>	<b>0.8</b>

Como colofón al estudio realizado se propone la estrategia que se muestra en el esquema de la figura 5.28, en base a los resultados obtenidos con los diferentes métodos. Dicho esquema representa una matización del diagrama general explicado en el capítulo dos y sintetizado en la figura 2.1. Volviendo sobre el esquema de la figura 5.28, en primer lugar se representa la captura de las dos imágenes involucradas en el proceso *IZ* (Imagen Izquierda) e *ID* (Imagen Derecha) con la configuración geométrica del sistema descrita en las secciones 5.1.1 y 4.2.2, que da lugar a la aplicación de la restricción de epipolaridad. A continuación se aplica el proceso de segmentación para excluir las texturas, tras el cual se determinan las propiedades (atributos) de los píxeles que se utilizan como características para la correspondencia. Este proceso se aplica en paralelo a ambas imágenes del par estereoscópico, que en la figura se ha identificado

como *EA*. La aplicación de las restricciones de epipolaridad y semejanza a través de los procesos correspondientes, identificados en el gráfico como *ES*, proporciona unos valores de similitud entre píxeles de una y otra imagen de suerte que gracias a la aplicación de la restricción de unicidad, naturalmente con los detalles especificados previamente, se decide sobre el mejor de los candidatos posibles, en este caso bajo el paradigma *MCDMF* por ser con el que mejores resultados hemos obtenido. Tras la aplicación de dicho método se obtiene un mapa de disparidad inicial, que es refinado posteriormente mediante el proceso *HNN*, con el que se obtiene el mapa de disparidad final.

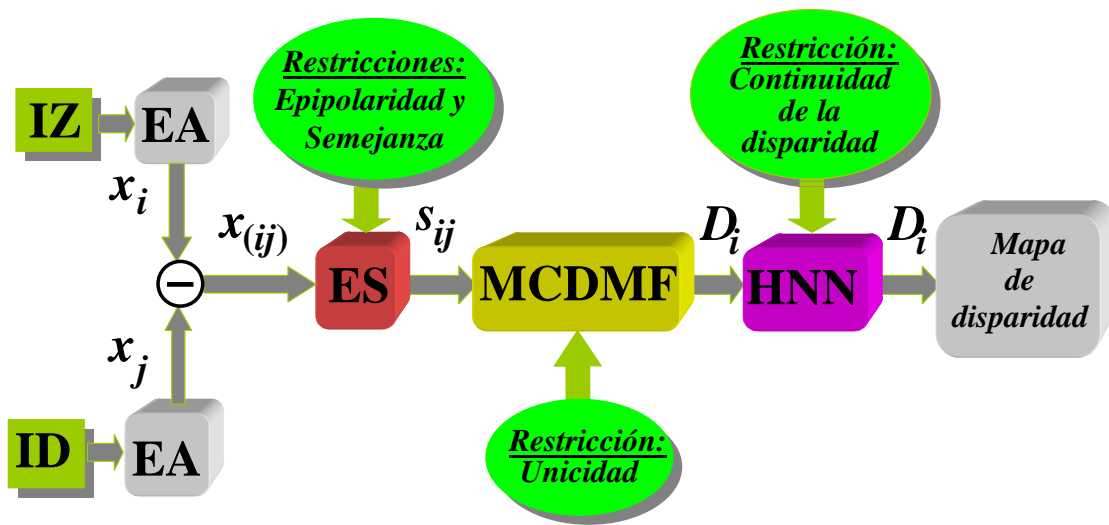


Figura 5.28 Propuesta de proceso de correspondencia estereoscópica basada en píxeles.

## 5.7 Resultados de la Correspondencia basada en regiones

Hasta ahora en las secciones precedentes se han analizado los resultados de la correspondencia basada en píxeles. Ahora se aborda el análisis de resultados de correspondencia mediante la segunda estrategia propuesta en este trabajo, es decir, la basada en regiones. Estos resultados se corresponden exactamente con los publicados en Herrera y col. (2009d).

En primer lugar y como paso previo a la aplicación del proceso de correspondencia, es necesario establecer los parámetros involucrados en el mismo. Con tal propósito se utiliza el conjunto de cuatro pares de imágenes estereoscópicas disponibles a tal efecto. De este modo es posible determinar los valores de los parámetros  $T_1$ ,  $T_2$  y  $T_3$  utilizados en la sección 4.3.3 y fijados a 0.3, 0.5 y 0.3

respectivamente; así como el valor umbral  $T_4$  de la sección 4.3.4, fijado a 6400 píxeles en ambos casos, de acuerdo con las razones esgrimidas en sendas secciones.

Procede analizar en esta etapa los resultados de los emparejamientos de regiones para las imágenes estereoscópicas sobre las que se ha aplicado el proceso de correspondencia. Para ello tomamos como referencia el par estereoscópico representado por las imágenes mostradas en las figuras 1.5(a) y (b). El punto de partida lo constituye las imágenes etiquetadas que aparecen mostradas en las figuras 5.10(a) y (b). Dadas las características de este tipo de imágenes, resulta relativamente fácil establecer las correspondencias sin más que utilizar la pericia basada en la observación humana.

En efecto, en la tabla 5.13 se recopilan los resultados de las correspondencias. La primera columna muestra el número de regiones etiquetadas (truncos) en la imagen izquierda, mientras que en la segunda columna aparecen las regiones que les corresponden, desde el punto de vista estereoscópico, en la imagen derecha según se desprende de la simple observación de ambas imágenes. Las tercera y cuarta columnas muestran las correspondencias individuales  $lk$  y  $rk$ , tal y como se describe en la sección 4.3.5. Por último, la quinta columna muestra la decisión final en términos de acierto (A) o fallo (F), siempre según la observación directa que se viene aplicando y teniendo como referencia las columnas primera y segunda.

Tabla 5.13 Resultados obtenidos empleando los momentos de Hu para establecer la correspondencia de regiones en ambos pares estereoscópicos.

Regiones imagen izquierda $L_i$	Regiones correspondientes Imagen derecha ( $R_j$ )	$lk$	$rk$	Decisión final correspondencia
1	5	7	7	A
2	1	7	7	A
3	3	7	7	A
4	2	7	7	A
5	4	7	7	A
6	sin correspondencia (oculto por 5)	0	0	A no se le asigna correspondencia
7	6	7	7	A
8	8	4	5	A
9	10	6	7	A
10	7	7	1	A
11	9	7	7	A
12	11	7	3	A

Tabla 5.13 *Cont.*

13	12	1	7	A
14	13	7	7	A
15	14	7	7	A
16	15	7	7	A
17	16	7	7	A
18	17	2	4	F
19	18	5	6	A
20	19	4	3	F

A partir de los resultados mostrados en la tabla 5.13, pueden inferirse las siguientes conclusiones:

1. Se observa que las regiones etiquetadas como 2, 1 y 3 en la imagen izquierda se corresponden con las regiones 1, 5 y 3 respectivamente en la imagen derecha. Sin la limitación de la restricción de orden posicional impuesta con respecto a las alturas y áreas de las regiones, sección 4.3.4, esta restricción sería claramente violada por la región 3 en la imagen izquierda porque el orden posicional con respecto a la pareja 1 y 5 no se preserva. En este caso la restricción de orden posicional se aplica solamente entre las regiones 2 y 1 en la imagen izquierda, y 1 y 5 en la derecha. Las alturas y áreas cumplen los requisitos especificados en la sección 4.3.4. Puede fácilmente deducirse que la restricción de orden posicional se viola para el caso de las regiones 19, 18 y 20 en la imagen izquierda, que se corresponden con las 18, 17 y 19, mientras que el orden correcto es 18, 19 y 17. Si se aplican los requisitos de la sección 4.3.4, dichos pares de regiones cumplen que los tamaños de sus áreas en píxeles no difieren más de un 33%, pero no el requisito de que las áreas deben superar el umbral  $T_4$ , es decir, en este caso la restricción de orden posicional no se aplica, siguiendo el diseño del procedimiento previsto.
2. *Oclusiones*: existe una clara oclusión que ha sido correctamente tratada. En efecto, la región 6 es perfectamente visible en la imagen izquierda, mientras que su homóloga en la imagen derecha se encuentra oculta por la 5. La estrategia propuesta en este trabajo de investigación no encuentra correspondencia para la mencionada región 6, como cabía esperar y por consiguiente, esta decisión se considera como un acierto.

3. *Ambigüedades*: existen dos tipos de ambigüedades que surgen dentro del área de interés en la parcela de muestra y fuera de la superficie ocupada por ésta. Al primer caso pertenece la ambigüedad entre la región 13 en la imagen izquierda y las regiones 12 y 7 en la imagen derecha. Al segundo caso pertenecen las regiones 18 y 20 en la imagen izquierda, donde ambas poseen como correspondencias preferidas las regiones 17 y 19. El primer caso se resuelve gracias a la aplicación de la restricción de orden posicional. Lamentablemente, en el segundo caso esta restricción no soluciona la ambigüedad, causando falsas correspondencias. Esta misma problemática se repite en otros pares de imágenes de prueba analizadas. A pesar de este contratiempo, la aplicación del método dentro del área de interés sigue siendo favorable, ya que las regiones donde se produce el fallo en la decisión se sitúan fuera de dicha área. Ciertamente esto podría suponer un problema en el futuro a la hora de ampliar la zona actual de exploración de 25m con sistemas de captura de imágenes más avanzados.
4. El porcentaje de correspondencias acertadas en el par estéreo mostrado en este trabajo es del 90%. En promedio, el porcentaje de éxito para los dieciséis pares de imágenes estereoscópicas analizados con características similares se sitúa en torno al 88.4%.

En general, la combinación de los dos procesos propuestos, segmentación y correspondencia, son adecuados y suficientemente robustos para el tipo de imágenes analizadas. En este sentido, su aplicación es válida para este tipo de imágenes captadas principalmente bajo las condiciones de iluminación descritas y con estructuras arbóreas de las características dadas en estas imágenes, entre las que destaca el hecho relevante de la escasez de hojas, además de la ya mencionada escasa luminosidad que se da por supuesta.

## 5.8 Resumen de los resultados obtenidos basados en regiones

A modo de resumen, en la figura 5.29 se muestra un esquema del proceso completo basado en regiones. En primer lugar se representa la captura de las dos imágenes involucradas en el proceso *IZ* (Imagen Izquierda) e *ID* (Imagen Derecha). A continuación se aplica el proceso de segmentación donde por un lado se realiza un



proceso de extracción regiones, que en nuestro caso constituyen las características bajo correspondencia, junto con el etiquetado de las mismas (*EC*). Una vez que todas las regiones han sido etiquetadas, para cada una de ellas extraemos los atributos necesarios (área, centroide y los siete momentos invariantes de Hu), tal y como se describe en la sección 3.3.3. Este proceso se identifica en la figura como *EA*. Tanto *EC* como *EA* se aplican a cada una de las imágenes del par. El centroide de cada región, considerado como atributo de la misma, se utiliza como elemento clave para su aplicación bajo la restricción epipolar, sirviendo como guía para la correspondencia, tal y como se explica en las secciones 5.1.1 y 4.3.2. Esta restricción se identifica en la figura como *EP*. Seguidamente, considerando el área de cada región en los términos descritos en la sección 4.3.3 bajo la restricción de semejanza (*SE*), se obtiene la lista de candidatos posibles para esa región de la imagen de referencia. A continuación, a partir de los momentos invariantes de Hu obtenidos de cada región, mediante la aplicación de la restricción de unicidad, se obtiene la región homóloga en la otra imagen del par estereoscópico. Si no se ha obtenido la unicidad buscada en la correspondencia, se produce una situación de ambigüedad, donde la aplicación simultánea de las restricciones de orden posicional, explicada en la sección 4.3.4, y unicidad resuelven este problema (*OU*), como se detalla en la sección 4.3.5. Tras la aplicación de dichas restricciones, obtenemos las disparidades de las regiones.

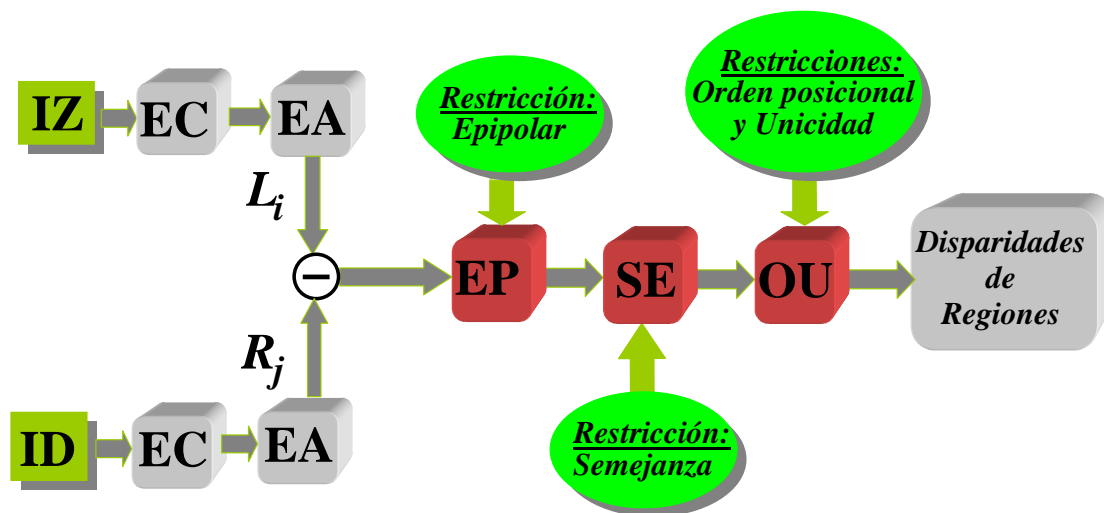


Figura 5.29 Propuesta de proceso de correspondencia estereoscópica basada en regiones.

## Conclusiones y Trabajo futuro

### 6.1 Introducción

Llegados a este punto se hace necesario realizar un balance global sobre las líneas de investigación abordadas y las conclusiones obtenidas. También analizaremos las perspectivas y posibilidades de futuro relacionadas con los temas de investigación abordados.

### 6.2 Conclusiones

La contribución principal de este trabajo de investigación radica en el diseño de una estrategia global para tratar de dar solución al problema de la correspondencia estereoscópica para un tipo característico de imágenes omnidireccionales obtenidas mediante un sistema óptico basado en las lentes conocidas como de “ojo de pez” y capturadas en entornos forestales.

Esta estrategia supone un paso fundamental y a la vez necesario para la consecución del objetivo último consistente en la obtención de ciertas medidas relacionadas con los entornos forestales de cara a su explotación y mantenimiento, tal y como se ha puesto de manifiesto en los capítulos anteriores. El planteamiento global de dicha estrategia contempla dos procesos bien diferenciados, donde los troncos de los árboles constituyen el objetivo fundamental de la misma. Los dos procesos mencionados son exactamente y en este orden, los siguientes:

- la segmentación de las imágenes del par estereoscópico
- la correspondencia entre características en ambas imágenes

La idea consiste en automatizar esos procesos hasta donde sea posible. Debido a la diversidad de imágenes que se pueden generar, los procesos anteriores van a depender mucho del entorno y de las condiciones lumínicas de las fotografías, siendo éste precisamente el caso que nos ocupa. En efecto, en el presente trabajo se han analizado imágenes procedentes de dos tipos de bosques bien diferenciados, lo que ha dado lugar a que los dos métodos que conforman la estrategia sean, a su vez, diferentes entre sí, y dependientes del entorno.

Uno de los entornos forestales se corresponde con bosques de pinares. Las imágenes procedentes del mismo se caracterizan por poseer una alta variabilidad en los niveles de intensidad. Esto es así debido a que su captura se realiza con una luminosidad intensa en días soleados. Dependiendo de la posición del sol en el momento de la captura, aparecen sobre la imagen tanto zonas con iluminación directa como difusa. Esto hace que idénticas estructuras en las imágenes, tales como los troncos de los árboles, puedan aparecer, y de hecho aparecen, bajo apariencias bien diferentes, lo que dificulta enormemente los correspondientes procesos.

Para solventar esta problemática, en el caso de los bosques de pinares, se ha aportado una solución que constituye una de las contribuciones más significativas de este trabajo. En efecto, en lo que respecta a la segmentación se realiza una propuesta orientada a la identificación de tres tipos de texturas que aparecen con profusión: texturas pertenecientes al cielo, al suelo y a las hojas de los pinos. El objetivo final consiste en su identificación para ser excluidas durante el posterior proceso de correspondencia. Bajo esta perspectiva se propone la utilización de dos clasificadores clásicos para la exclusión de texturas procedentes del cielo y del suelo, a saber: el *Agrupamiento Borroso (FC)* y el *estimador paramétrico Bayesiano (PB)*. Para la identificación de las hojas de los árboles se propone un método típico de identificación de texturas con alta variabilidad en los niveles de intensidad en un área limitada por una ventana de vecindad, habiendo elegido la *varianza* como descriptor de este tipo de texturas por su simplicidad y efectividad. Los experimentos realizados han probado ambos métodos de clasificación, tanto en su versión individualizada como mediante la

combinación de ambos. Los resultados obtenidos nos sugieren que es bajo su versión combinada donde se mejoran los resultados.

Debido a la problemática presentada por este tipo de imágenes, el posterior proceso de correspondencia tras la segmentación se realiza píxel a píxel, toda vez que se han excluido de las imágenes los tres tipos de texturas carentes de interés. Para tal fin se han propuesto y probado una serie de estrategias encaminadas al emparejamiento de los píxeles entre ambas imágenes del par estereoscópico. Todas ellas se fundamentan en la aplicación de las restricciones habitualmente utilizadas en correspondencia estereoscópica, a saber: *epipolar*, *semejanza* y *unicidad*. La decisión final sobre los emparejamientos se lleva a cabo mediante la adaptación al problema de las siguientes tres estrategias procedentes del ámbito de la combinación de clasificadores: *Integral Fuzzy de Choquet (CFI)*, *Integral Fuzzy de Sugeno (SFI)*, *Teoría de Dempster-Shafer (DES)*. A estas tres se les añade una cuarta cuyo fundamento se centra en la *Toma de Decisiones Multicriterio* bajo el paradigma de la lógica *Fuzzy (MCDMF)*.

Como atributos para la toma de decisiones, por medio de cualquiera de los cuatro métodos previos, se han utilizado dos tipos de enfoque. En primer lugar dos aproximaciones encuadradas dentro de lo que se conoce como *basadas en el área* (coeficiente de correlación y textura) y en segundo lugar cuatro enmarcadas dentro de lo que se conoce como *basadas en las características* (color en el espacio *RGB*, módulo del gradiente, ángulo del gradiente y Laplaciana), haciendo un total de seis atributos. Los métodos anteriores poseen la capacidad de combinar dichos atributos de una manera adecuada. Los resultados de la combinación generados por cada uno de los métodos se han comparado entre sí e incluso con otros métodos clásicos combinados. Además, los resultados de las estrategias combinadas se han comparado con los resultados obtenidos mediante la aplicación de cada atributo de forma individualizada. En cualquier caso, las estrategias propuestas han resultado ventajosas frente a las individuales, resultando la *MCDMF* como la más favorable de todas ellas para el conjunto de imágenes analizado.

El objetivo de las anteriores estrategias es la obtención de un mapa de disparidad que proporciona la diferencia en grados entre los píxeles homólogos de ambas imágenes, principalmente para el caso de los troncos. Tras la obtención de este mapa de disparidad, el siguiente paso ha sido la aplicación de un método orientado a mejorar

dicho mapa, principalmente desde el punto de vista de su suavizado mediante la homogeneización de zonas con una irregular distribución de los niveles de disparidad. Para realizar este suavizado se proponen dos nuevas estrategias especialmente diseñadas para este propósito. Se trata por un lado del paradigma de los *Mapas Cognitivos Fuzzy (FCMs)* y por otro de la *Red Neuronal de Hopfield (HNN)*. En ambos casos, su fundamento se basa en la aplicación de la restricción de *continuidad de la disparidad*, inspirada en los principios de proximidad y similitud de la escuela de psicología de la Gestalt. Según estos principios se preserva la misma relación espacial entre un píxel y sus vecinos en las imágenes, que la existente en el espacio tridimensional. La propuesta basada en la *HNN* produce mejores resultados que los obtenidos mediante los *FCMs*.

El segundo entorno forestal analizado es el que se corresponde con las imágenes de robles, cuya característica fundamental es su baja luminosidad. Los procedimientos de segmentación y correspondencia se han diseñado aprovechando y explotando esta circunstancia, diferenciándose, por este motivo, sustancialmente de los correspondientes procesos diseñados para el entorno forestal de los pinos.

Desde el punto de vista de la segmentación se aplican varias técnicas de procesamiento de imágenes para la extracción de las regiones correspondientes a los troncos de los robles. En este enfoque son las regiones los elementos utilizados como características, a diferencia del anterior cuya base son los píxeles. De las regiones se extraen una serie de atributos (área, centroide y los siete momentos invariantes de Hu) que permiten establecer las correspondencias por aplicación de las siguientes restricciones estereoscópicas: *epipolar*, *semejanza*, *orden posicional* y *unicidad*.

Se trata de una estrategia totalmente novedosa en este campo de aplicación que consigue unos resultados muy significativos cuando se comparan con las correspondencias visuales obtenidas por el experto humano. Con ella se establece la base, que constituye el punto de partida para la obtención de las medidas previstas, algunas de ellas mencionadas en el capítulo uno, para la elaboración de los inventarios forestales.

## 6.3 Trabajo futuro

A la vista de los resultados obtenidos y teniendo en cuenta la experiencia acumulada con los conjuntos de imágenes tratadas durante el desarrollo de este trabajo de investigación, vamos a proponer una serie de consideraciones para establecer las pautas a seguir en las futuras líneas de investigación. Distinguimos entre aquellas derivadas del propio sistema de medición y aquellas otras provenientes de los dos enfoques utilizados, es decir, el basado en píxeles y el que utiliza regiones. Desde la perspectiva de estos dos enfoques se contemplan los aspectos relacionados con sendos procesos de segmentación y correspondencia.

### 6.3.1 Sistema de medición estereoscópico

Una de las propuestas que se formulan respecto del sistema físico de medición consiste en el equipamiento del mismo con un par de cámaras de disparo simultáneo, en lugar de la única cámara con la que está dotado el sistema actual. De esta forma se pueden evitar los inconvenientes surgidos durante el desplazamiento de la única cámara entre las posiciones de la línea base. Entre estos inconvenientes se puede citar el hecho de que entre la captura de una imagen y la captura de la segunda imagen del par transcurre un cierto intervalo de tiempo, durante el cual puede variar la escena o las condiciones de iluminación en la misma. Por ejemplo, una ráfaga de viento puede provocar ligeros desplazamientos posicionales de las ramas, principalmente en la parte superior, o un cambio de iluminación de la escena provocado por el paso transitorio de una nube o la apertura de un claro, puede originar variaciones sustanciales en la iluminación de un mismo tronco en sendas imágenes del par, dificultando las tareas de correspondencia.

Además de lo anterior, cabe mencionar que si bien el sistema se fija con un mecanismo de anclaje al suelo, a veces durante el desplazamiento manual de la cámara se pueden introducir perturbaciones en el posicionamiento, que sin duda van a introducir errores que afectan a las variables geométricas del sistema y en consecuencia al proceso de correspondencia, que tiene en cuenta la geometría principalmente a través de la restricción epipolar.

### 6.3.2 Procedimiento basado en píxeles

Desde la perspectiva del enfoque basado en píxeles, con vistas a su aplicación en entornos caracterizados por las diferencias de iluminación, se proponen las siguientes líneas de actuación futuras:

1. Perfeccionar los procesos de segmentación y correspondencia, a la vez que se extiende su ámbito de aplicación a otros entornos forestales en la línea indicada en el apéndice B, donde las imágenes presentan características muy diferentes.
2. Completar el proceso global de la visión estereoscópica mediante la implementación del módulo de cómputo de distancias, que aparece como “obtención de la estructura 3D en la escena” en la figura 2.1. En una primera fase la pretensión consiste en realizarlo sobre un entorno simulado para posteriormente aplicarlo a los entornos forestales reales. El entorno simulado, creado con el software libre de evaluación proporcionado por Truevision3D (2010), permite conocer con exactitud las distancias de los árboles, además de permitir recrear distintos tipos de texturas, imitando a las propias texturas naturales. En la figura 6.1 se muestra una imagen representativa de este tipo de escenarios simulados, actualmente en proceso de desarrollo.



Figura 6.1 Par estereoscópico generado con Truevision3D (a) imagen izquierda; (b) imagen derecha.

3. Realizar un análisis más en profundidad relativo al estudio del efecto y consecuencias de tomar las imágenes de una misma escena con distintas

medidas de posicionamiento de la cámara sobre su raíl de deslizamiento, esto es, distintas medidas de la línea base.

4. Dado que la combinación de clasificadores se perfila como una estrategia que mejora los resultados de los clasificadores individuales *FC* y *PB*, respecto de la exclusión de texturas no deseadas, tal y como se ha explicado en la sección 5.2.3, una propuesta de mejora sería la utilización de otros métodos de combinación tales como los propuestos en Guijarro (2009), donde además de la combinación se proponen otros clasificadores individuales que pueden producir resultados más relevantes. En cualquier caso se hace necesario insistir más en la búsqueda de métodos de clasificación tanto en lo que se refiere a la mejora de los ya existentes como en lo relativo a la búsqueda de nuevas alternativas de segmentación tal y como se apunta en el apéndice B.
5. Sobre la base del suavizado del mapa de disparidad, se podrían utilizar algunos métodos de optimización, tal como el enfriamiento simulado (“*simulated annealing*”), donde también la restricción de continuidad de la disparidad constituiría su fundamento. Los principios de la Gestalt serían igualmente aplicables, con la ventaja de que este proceso está gobernado por una función de energía que debe minimizarse. Existen algunos estudios en la literatura concluyentes sobre las ventajas de la técnica basada en enfriamiento simulado frente a la técnica *HNN*; así en Geman y Geman (1984) o Haykin (1994), se afirma que cuando la temperatura, involucrada en el proceso de enfriamiento, satisface determinadas restricciones, entre las que se encuentran el hecho de comenzar con una temperatura inicial alta y el diseño de un proceso de enfriamiento relativamente lento, el proceso converge hacia mínimos globales, lo cual puede ser una garantía de estabilidad y por tanto de mejora de resultados.
6. Cuando se obtienen los mapas de disparidad de los troncos, en el caso del procesamiento por píxeles, habría que realizar un proceso de identificación de la región que caracteriza el tronco mediante un crecimiento de regiones tomando como referencia los valores de disparidad en lugar de las intensidades. De esta forma, se obtendrían agrupaciones de disparidades que se identificarían como regiones completas, constituyendo el paso previo para la obtención de las medidas previstas de cara al objetivo final de los inventarios forestales. Esto



sería equivalente a la construcción de los denominados Mapas Digitales del Terreno (“*Digital Elevation Models*”, *DEM*) en los procesos estereoscópicos convencionales.

7. Los enfoques combinados propuestos en este trabajo (*CFI*, *SFI*, *DES*, *MCDMF*) pueden aplicarse a cualquier sistema que requiera correspondencia estereoscópica, y en consecuencia a sistemas estereoscópicos convencionales con ópticas basadas en proyección de perspectiva y ejes ópticos paralelos o convergentes. La única diferencia afectaría básicamente a la aplicación de la restricción epipolar, ya que las líneas epipolares en estos sistemas son líneas rectas, en lugar de las líneas curvadas que caracterizan nuestro sistema basado en la geometría definida en el capítulo cuatro.
8. Finalmente, aunque la estrategia propuesta para resolver el problema de la correspondencia estereoscópica, basada en sistemas con lentes de ojo de pez, se ha desarrollado con vistas a la realización de inventarios forestales, las técnicas involucradas en la misma podrían ser fácilmente adaptadas para navegación autónoma de vehículos en entornos no estructurados de exterior, donde las imágenes presentan características y comportamientos similares a los de las imágenes analizadas en este trabajo. De hecho, hace ya más de una década se realizaron experimentos en este sentido utilizando sistemas estereoscópicos cuya óptica está basada en las lentes de ojo de pez (Shah y Aggarwal, 1997), y aunque bien es cierto que no suele ser habitual la utilización de estos sistemas para navegación autónoma, hasta donde hemos podido investigar las técnicas planteadas en nuestra investigación pueden ser fácilmente adaptables para su incorporación a sistemas estereoscópicos convencionales. En esta línea hemos iniciado ya la investigación al amparo del proyecto titulado *Sistema de Vigilancia, Búsqueda y Rescate en el Mar mediante colaboración de Vehículos Autónomos Marinos y Aéreos*, perteneciente al Plan Nacional de I+D+i con número de referencia DPI2009-14552-C02-01. En este proyecto se contempla como un objetivo importante la localización de los vehículos involucrados mediante distintos sistemas combinados, entre los que se incluye la visión estereoscópica. En principio, las primeras pruebas experimentales se han comenzado ya, si bien de momento en un entorno simulado, como paso previo a

los entornos estructurados de interior, a través del vehículo mostrado en la figura 6.2, y de aquí a los entornos no estructurados de exterior.



Figura 6.2 Robot equipado con sistema de visión estereoscópica convencional para navegación autónoma.

### 6.3.3 Procedimiento basado en regiones

Respecto de las futuras consideraciones sobre los procesos relativos al enfoque basado en regiones cabe mencionar:

1. Durante el proceso de segmentación relativo a las imágenes de robles, un problema importante que se plantea deriva del hecho de que en la parte inferior de los troncos, éstos se confunden con el suelo en las imágenes, haciendo muy difícil su identificación. Pues bien, cabe la posibilidad de diseñar nuevos procesos que consideren sólo la parte diferenciable con respecto al cielo y de ahí extraer información relevante sin necesidad de tener que discriminar entre suelo y tronco en la parte inferior. Esto debe llevarse a cabo necesariamente en colaboración con los técnicos forestales, quienes han de determinar la validez de este tipo de aproximaciones.
2. Tal y como se muestra en la figura 1.6, las imágenes son en realidad en color, habiéndose obtenido por realzado mediante igualación de histograma de la componente de intensidad del modelo *HSI*. Pues bien, cabe la posibilidad de investigar sobre nuevas técnicas de realzado con el fin de tratar de conseguir una

mayor diferenciación de las partes actualmente indistinguibles (Pajares y Cruz, 2007a; Pratt, 1991).

3. En la misma línea de lo expresado en el punto anterior, no sólo se pueden investigar los aspectos relacionados con el realzado sino los de cualquier otro tipo de procesamiento dirigido en este sentido, como estudiar la representación de estas imágenes en otros modelos de color, tales como *HSI*, *YIQ*, *CMYK* entre otros (Pajares y Cruz, 2007a).



# Sistemas de Visión Estereoscópica

## A.1 Introducción

La visión del Hombre es binocular. Esto quiere decir que el cerebro humano interpreta la realidad a partir de dos imágenes con pequeñas diferencias entre sí, debidas a la separación física existente entre los ojos, aproximadamente de unos  $65mm$  (Howard y Rogers, 1995). Esta separación hace que un mismo punto de la escena tridimensional (3D) se proyecte en posiciones relativas diferentes en las retinas de ambos ojos. En efecto, si hubiese posibilidad de superponer ambas retinas, las imágenes de un punto aparecerían separadas entre sí una cierta distancia conocida como *disparidad* o *paralaje*, que es utilizada por el cerebro para percibir la profundidad a la que se encuentra el punto en la escena 3D. En la figura A.1(a) se muestra un esquema de un sistema de visión estereoscópica biológico, que observa dos objetos (estrella y triángulo) a distintos niveles de profundidad en la escena 3D. En las correspondientes retinas de los dos ojos se obtienen las respectivas imágenes. Si solapamos ambas retinas obtenemos la imagen de la figura A.1(b), en la que se observa que la separación relativa entre las imágenes de los dos triángulos es menor que la separación relativa entre las imágenes de las estrellas. Este fenómeno se explica por el hecho de que la estrella en la escena 3D se encuentra más próxima a los ojos que el triángulo. Estas separaciones relativas entre las imágenes de un mismo objeto es lo que previamente se ha denominado *disparidad*, siendo como hemos dicho la base de la visión estereoscópica. Si bien el cerebro dispone de otras

señales para percibir la tridimensionalidad del entorno, como la perspectiva, la interposición o solapamiento, y el enfoque, la disparidad es probablemente la más importante.

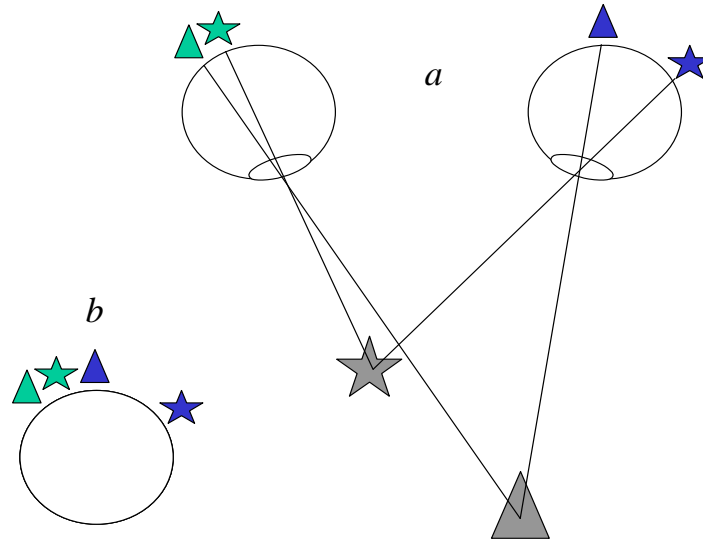


Figura A.1 (a) Sistema de visión estereoscópica biológico, con dos objetos en la escena 3D (estrella y triángulo) a distintos niveles de profundidad; (b) Superposición de las imágenes de ambos ojos, donde se observa que la separación relativa entre los triángulos es menor que entre las estrellas, lo que está en relación inversa con la distancia de los correspondientes objetos en la escena 3D.

El desarrollo de la fotografía durante el siglo XIX permitió la aparición de las primeras cámaras y visores estereoscópicos. Durante el siglo XX, el desarrollo de la informática ha revolucionado esta técnica, con aplicaciones en cartografía, geología, robótica por citar sólo algunos ejemplos.

En este apéndice se analizan las posibilidades de la visión estereoscópica y se describen los distintos sistemas de generación de imágenes estereoscópicas, valorando para cada uno de ellos su utilidad y limitaciones. Este análisis se completa con una descripción de las áreas donde las técnicas de visión estereoscópica pueden realizar las aportaciones más importantes, según el desarrollo actual de la técnica.

## A.2 Visión estereográfica y señales de profundidad

La evolución ha hecho que todas las especies animales que precisan de una percepción precisa de la profundidad espacial, particularmente los depredadores, posean visión binocular estereoscópica, al solaparse en un alto porcentaje el ángulo de visión de ambos ojos.

La disparidad retinal no es la única señal de profundidad utilizada por el cerebro para percibir la tridimensionalidad del entorno. Otras señales permiten interpretar adecuadamente la profundidad de una imagen plana, algo muy útil para individuos con dificultades severas de visión en uno de los ojos:

- **Perspectiva cónica:** el hecho de que los objetos disminuyan aparentemente de tamaño al alejarse o que las líneas paralelas converjan constituyen señales de profundidad básicas (figura A.2).
- **Interposición o solapamiento:** los objetos próximos ocultan a los objetos más alejados.
- **Iluminación y sombreado:** la tridimensionalidad de los objetos se ve realzada por medio de la iluminación lateral, pues se intensifican los efectos de luz y sombra.
- **Atenuación y desplazamiento cromático al azul:** los objetos lejanos, aparecen más difuminados. También se observa un desplazamiento cromático hacia los tonos azules en los objetos más alejados.
- **Enfoque:** para percibir nítidamente el objeto observado, el ojo debe adaptarse, teniendo en cuenta la distancia a la que se encuentre.

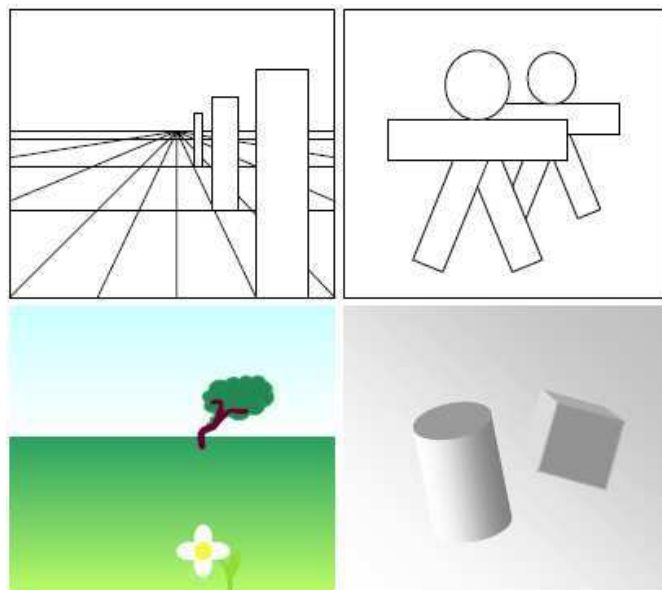


Figura A.2 Señales de profundidad: perspectiva, solapamiento, desviación cromática y sombreado.

### A.3 Creación y visualización de imágenes estereoscópicas

Históricamente, se atribuye al físico Charles Wheatstone (1802–1875) la interpretación de los fundamentos de la visión estereoscópica, en el siglo XIX, si bien existen antecedentes que se remontan al siglo XVI e incluso antes. El desarrollo de la **fotografía** durante el siglo XIX conduce a la aparición de las primeras cámaras estereoscópicas, y como consecuencia de ello a los primeros visores estereoscópicos. En estos visores las fotos que constituyen el par estereográfico son montadas una junto a la otra (en un cartón si son positivos o en un cristal si se trata de diapositivas). El visor consta de unas lentes que facilitan la acomodación de la vista (enfoque) y un bastidor que garantiza que cada ojo perciba únicamente la imagen que le corresponde, figura A.3.



Figura A.3 Visor estereoscópico de finales del siglo XIX.

Por tanto, el fundamento de estos sistemas se basa en la creación y visualización de imágenes estereoscópicas: captar o crear dos imágenes de un mismo entorno a partir de dos puntos de vista diferentes, para luego conseguir, mediante distintos sistemas, que cada ojo perciba únicamente la imagen que le sea asignada. La posibilidad de generar imágenes mediante el uso de computadores abre nuevas posibilidades a los sistemas estereoscópicos, anteriormente limitados a la representación de fotografías.

Existe una gran diversidad de sistemas de visualización estereoscópica, algunos de los más importantes se describen brevemente a continuación (Lipton, 1997):

- **Visión paralela:** es el primer y más primitivo sistema de visualización. El observador debe centrar su atención en un punto en el infinito, manteniendo los ejes de visión paralelos. Las imágenes no pueden tener un tamaño superior a  $65mm$  entre sus centros. Para facilitar la visión paralela se utilizan lentes con el fin de acomodar la vista y evitar que cada ojo perciba lateralmente la imagen correspondiente al otro ojo.
- **Visión cruzada:** en este caso las líneas de visión de ambos ojos se cruzan, fijándose en un punto medio más cercano. De esta forma el ojo derecho observa la imagen izquierda y viceversa. Es posible emplear imágenes con formatos mayores de  $65mm$ .
- **Anaglifos:** las imágenes que forman el par estereográfico se representan superpuestas, previamente coloreadas utilizando tonos complementarios (rojo – azul, rojo – verde, ámbar – azul). El resultado puede ser una imagen en color o en tonos de escala de gris. En la figura A.4 se muestran sendos anaglifos correspondientes a los pares de imágenes estereoscópicas mostrados en las figuras 1.3 y 1.5, respectivamente. Como puede observarse, las tonalidades azules provienen de la imagen izquierda, mientras que las tonalidades rojas proceden de la derecha. La observación de estas imágenes mediante unas gafas apropiadas con el filtro rojo en el cristal izquierdo y el azul en el derecho, permite una visualización tridimensional de la escena. Un ejemplo de este tipo de gafas puede observarse en la figura A.5.

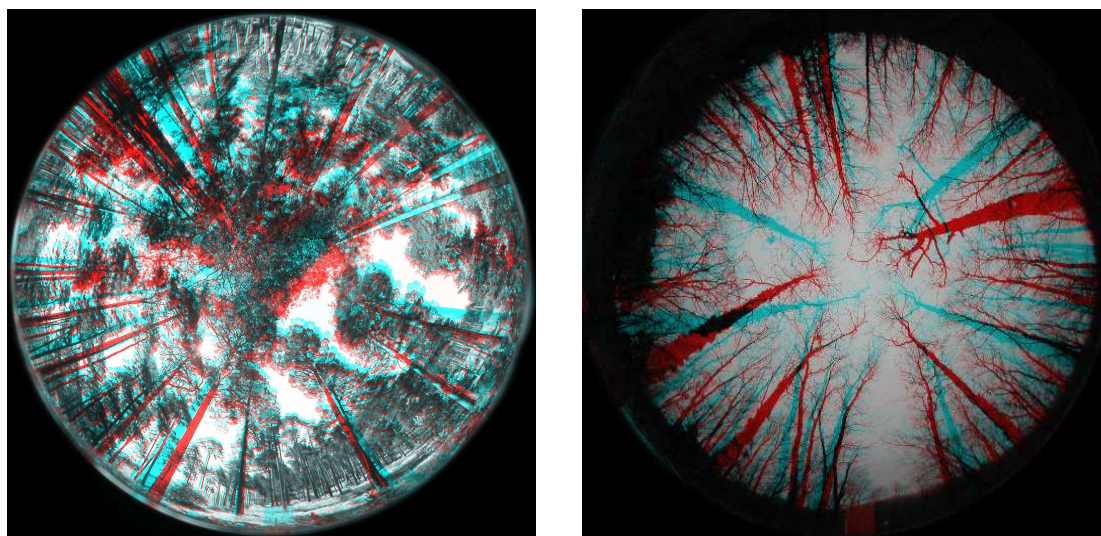


Figura A.4 Anaglifos correspondientes a dos pares estereoscópicos analizados.





Figura A.5 Gafas para visualización de Anaglifos.

- **Polarización:** las dos imágenes son proyectadas mediante sendos dispositivos dotados de filtros polarizadores, girados  $90^\circ$  uno respecto del otro. La pantalla debe conservar la polarización de la luz reflejada. La figura A.6 muestra un ejemplo del polarizador de la empresa Dimap (2010). El observador debe utilizar unas gafas con lentes polarizadas de forma congruente con los proyectores, figura A.7(a).



Figura A.6 Ejemplo del polarizador de la empresa Dimap.

- **Obturación** (*shutter glasses*): en este caso las dos imágenes son mostradas por un monitor de forma alternativa, con una frecuencia de refresco suficiente. El observador debe utilizar unas gafas dotadas de dos lentes de cristal líquido

(LCD) que se oscurezcan alternativamente de forma sincronizada con el monitor, figura A.7(b).

- **Cascos de realidad virtual** (*Head Mounted Display, HMD*). En este caso, el observador dispone de un casco especial con dos monitores miniaturizados, donde se proyectan las imágenes del par estereoscópico.
- **Monitor lenticular** o *auto-estéreo*. El monitor proyecta una imagen tridimensional sin necesidad de que el observador utilice ningún otro dispositivo. Para ello se adhiere un filtro lenticular sobre la pantalla. Por un simple efecto de difracción, el filtro lenticular ofrece a cada ojo la imagen deseada, ocultando la otra. Este sistema también puede ser utilizado con imágenes impresas.



Figura A.7 Ejemplos de sistemas para visualización estereoscópica:  
(a) lentes polarizadas; (b) gafas obturadoras.

## A.4 Áreas de aplicación de los sistemas de visión estéreo

Es evidente, que una de las áreas es la presentada a lo largo de este trabajo de investigación. No obstante, los sistemas de visión estereoscópica ofrecen al usuario otras posibilidades de aplicación en diferentes campos, que se pueden agrupar en cuatro grandes bloques, a saber:

- representación de información gráfica compleja
- telepresencia
- realidad virtual
- entrenamiento de la percepción espacial

### A.4.1 Representación de información gráfica compleja

Dentro de este grupo de utilidades destacamos el uso de las representaciones estereoscópicas en *Geología*, *Química* y *Medicina*, si bien la lista de áreas se podría extender bastante más.

La Fotogeología es la rama de la **Geología** encargada de estudiar los sucesos geológicos a través de fotografías aéreas estereoscópicas (Vergara, 1988). Mediante la composición de pares estereoscópicos, el experto identifica diferentes rasgos geológicos en el terreno, imperceptibles en fotografía bidimensional (figura A.8).

En **Química** es de gran utilidad la visualización de moléculas complejas mediante técnicas estereoscópicas. En **Medicina** la estereoscopía es utilizada como herramienta didáctica en la interpretación de imágenes para el diagnóstico. Así, se obtienen imágenes estereoscópicas a partir de radiografías, ecografías, tomografía (TAC) y resonancia magnética nuclear (RMN).



Figura A.8 Visualización de fotografías aéreas estereoscópicas en Fotogeología, mediante visión paralela asistida con lentes.

### A.4.2 Telepresencia

El número de aplicaciones orientadas a facilitar el trabajo en entornos hostiles, peligrosos o simplemente difíciles se incrementa continuamente. En general se emplean robots teledirigidos, dotados de cámaras estereoscópicas. El operario, mediante gafas obturadoras o cascos de inmersión, recibe la imagen captada por esas cámaras. Las aplicaciones son muy variadas, destacando entre otras, la desactivación de explosivos, la exploración espacial, la exploración submarina o las intervenciones quirúrgicas por medio de instrumental robótico.

### A.4.3 Realidad virtual

La creación de imágenes tridimensionales mediante computador adquiere una nueva utilidad cuando se combina con las técnicas estereoscópicas.

Este enfoque es útil en la fase de diseño en áreas tales como la arquitectura o la industria del automóvil. También es útil para el entrenamiento, mediante simuladores, del manejo de herramientas y vehículos (vuelo, conducción, navegación, etc.) Se ha aplicado como herramienta de conocimiento y divulgación del patrimonio histórico artístico (Zheng, 2000). Finalmente, la realidad virtual facilita la *didáctica* de un sinnúmero de materias, desde la medicina hasta el arte.

### A.4.4 Entrenamiento de la percepción espacial

Finalmente, el entrenamiento de la capacidad perceptual espacial humana resulta especialmente valorado en las carreras técnicas. En este caso las técnicas de visión estereoscópica son utilizadas para entrenar la capacidad del alumno a representar en su imaginación una realidad tridimensional a partir de su representación plana y viceversa.

La capacidad de realizar esa transición es considerada básica, al menos en la formación de los ingenieros y arquitectos. El dominio de los sistemas de representación es tan importante para el técnico como la capacidad matemática. Sin embargo, un número importante de alumnos carecen de esa habilidad, debido a que la representación tridimensional en entornos artificiales no resulta una tarea fácil. Ello hace especialmente interesante y útil el desarrollo de aplicaciones didácticas que utilicen las técnicas de visión estereoscópica como herramienta de entrenamiento.



# B

## Segmentación y correspondencia: nuevos retos

### B.1 Ampliación de la investigación: líneas de actuación

A lo largo de la presente memoria se han planteado dos estrategias aplicables a dos tipos diferentes de imágenes. En ambos casos se ha puesto de manifiesto que los procesos de segmentación y correspondencia, involucrados en las mismas, constituyen dos procesos inseparables.

Se ha pretendido diseñar procesos con una validez lo más general posible, de forma que sean aplicables al mayor número de tipos diferentes de imágenes procedentes de otros bosques.

Siguiendo esta línea de investigación hemos tratado de aplicar las técnicas diseñadas a un nuevo tipo de imágenes estereoscópicas procedentes de bosques de robles, en concreto a roble rebollo (*Quercus pyrenaica* Willd.). En la figura B.1(a) y (b) se muestra un par representativo de imágenes. Fácilmente se deduce que se trata de un tipo de imágenes que podría asimilarse, por sus características de intensidad luminosa, a las procedentes del bosque de pinos analizadas como primer tipo de imágenes en este trabajo de investigación. Este hecho sugiere la aplicación de la misma estrategia diseñada para aquel tipo de imágenes.

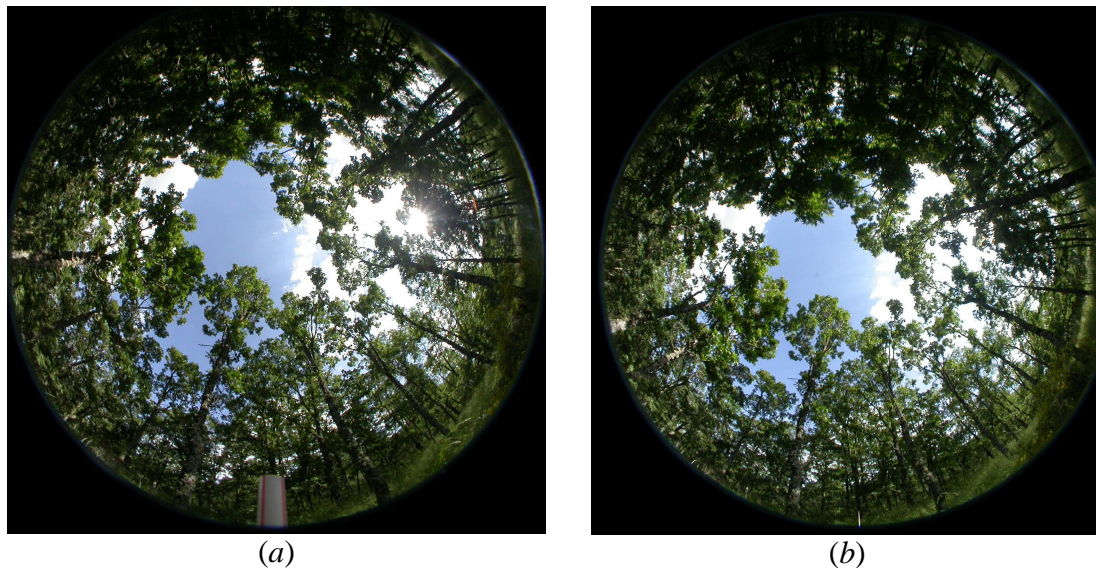


Figura B.1 Par de imágenes estereoscópicas omnidireccionales procedentes de robles:  
(a) izquierda; (b) derecha

El primer paso consiste en la segmentación de las imágenes, según el proceso descrito en el capítulo tres. No obstante en este caso surgen dos importantes dificultades. La primera procede del hecho de que el espesor de las hojas de los árboles no permite la aplicación de la técnica utilizada en los pinos para discriminar dichas texturas en función de la varianza de los niveles de intensidad, tal y como se describe en la sección 3.2.2.

Podría pensarse en prescindir de la parte relativa a la identificación de las texturas de las hojas y englobar su identificación dentro del proceso de clasificación de la hierba del suelo, ya que en el modelo de color *RGB* las texturas de las hojas y de la hierba presentan firmas espectrales relativamente similares. Cuando se aplica esta idea surge de nuevo un problema, y es que gran parte de los troncos llega a clasificarse como texturas a excluir por su pertenencia a la misma clase que la de las hierbas. Este hecho aflora relativamente cuando sobre las imágenes originales de la figura B.1 se aplica una transformación de color desde el modelo *RGB* al *HSI* y posteriormente, sobre la componente de intensidad *I* se aplica un realzado mediante ecualización uniforme de histograma (Pajares y Cruz, 2007a); el resultado es el que se muestra en las imágenes de la figura B.2(a) y (b) respectivamente. Esto conduce a que muchos píxeles en los troncos y en sus proximidades sean erróneamente clasificados. En efecto, aplicando el proceso de segmentación bajo la suposición de identificar sólo dos tipos de texturas a excluir (cielo y hojas junto con la hierba del suelo), el resultado de la segmentación es el que se muestra en la figura B.3(c) sobre el recorte de la imagen B.3(a). Partiendo de



este resultado obtenido mediante segmentación se aplica el proceso de correspondencia descrito en la sección 4.2 bajo el paradigma *MCDMF* y seguido del suavizado del mapa de disparidad basado en el proceso *HNN*, elegidos ambos por ser los que mejores resultados produjeron en su momento, según se refleja en el análisis del capítulo cinco. En la figura B.3(b) se muestra el recorte homólogo de la imagen derecha. El mapa de disparidad así obtenido es el que se muestra en la imagen de la figura B.3(d).

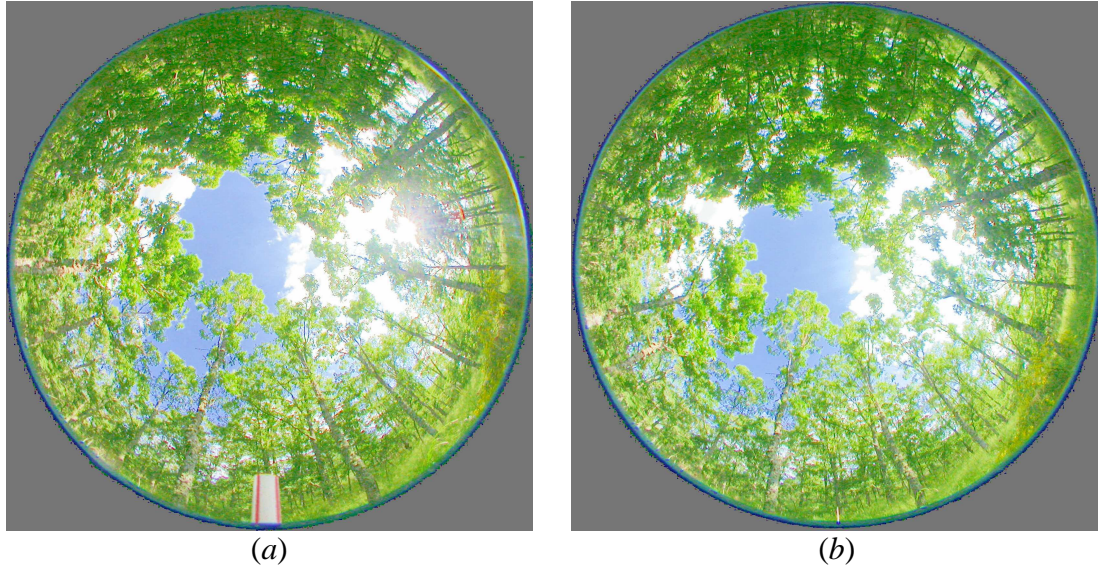


Figura B.2 Imágenes realizadas a partir de las mostradas en la figura B1 mediante igualación de histograma sobre la componente de intensidad en el modelo de color *HSI*.

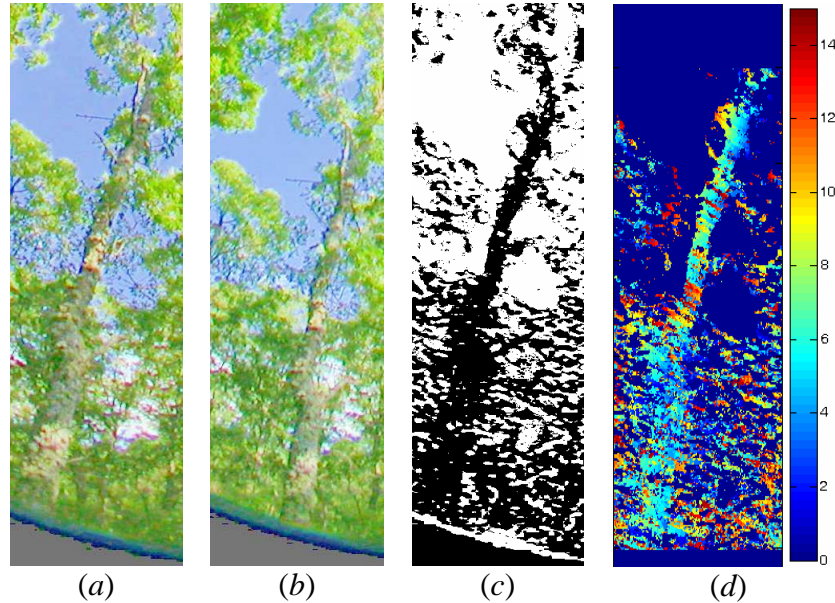


Figura B.3 (a) y (b) Recortes homólogos procedentes de las imágenes estereoscópicas B.2(a) y (b) respectivamente; (c) resultado del proceso de segmentación considerando sólo texturas de cielo y hojas junto con hierba del suelo; (d) disparidades obtenidas con el rango de variación mostrado en la barra de color.



Como se puede observar a partir del mapa de disparidad obtenido, existen muchos píxeles clasificados erróneamente. Para verificar la procedencia del problema, realizamos una segmentación manual sobre el resultado de la segmentación mostrado en B.3(c) hasta conseguir la imagen binaria de la figura B.4(a), con ella el mapa de disparidad obtenido es el mostrado en la imagen de la figura B.4(b), sobre la que puede apreciarse claramente la mejoría obtenida con respecto al obtenido en la figura B3(d).

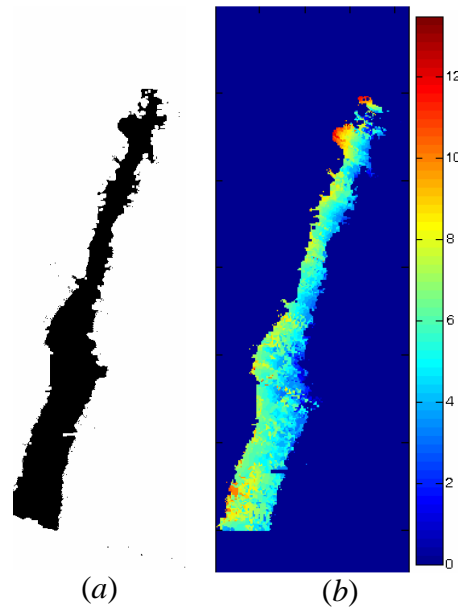


Figura B.4 (a) Resultado de la segmentación manual; (b) disparidades obtenidas con la variación mostrada en la barra de color.

## B.2 Resumen del estudio

Los experimentos anteriores ponen de manifiesto una serie de hechos relevantes que han de tenerse en cuenta a la hora de ampliar el campo de aplicación de los procedimientos propuestos en el presente trabajo de investigación:

- 1) El proceso de segmentación es un elemento fundamental del proceso global de correspondencia que requiere un esfuerzo importante. La clasificación de texturas resulta ser un aspecto clave del mismo. Esto corrobora que ambos, segmentación y correspondencia, han de considerarse como una entidad única.
- 2) Probablemente, en lugar de utilizar el modelo de color *RGB*, podría pensarse en utilizar mecanismos de clasificación e identificación de texturas basados en el modelo *HSI*, ya que en este tipo de imágenes la

componente  $S$  de saturación exhibe propiedades de interés. En este sentido se han realizado algunos experimentos basados en este modelo, aunque sin resultados satisfactorios hasta el momento ni a nivel de segmentación ni de correspondencia, lo que corrobora la necesidad de profundizar en la investigación. En la figura B.5(a) se muestra el resultado de la transformación de la imagen de la figura B.1(a) al modelo de color  $HSI$  (Pajares y Cruz, 2007a). En ella se observa la tendencia a la saturación de las texturas que han de ser excluidas.

- 3) Por otra parte, podría pensarse en la aplicación de otras técnicas de segmentación específicas en los casos en los que las texturas a excluir son de dos tipos. Una de ellas es la que se basa en el método de segmentación de regiones de Otsu (1979) para separar las texturas pertenecientes al cielo de la imagen  $H$  en el modelo de color  $HSI$ , tras lo cual se separaría la hierba y las hojas, de nuevo mediante Otsu, pero esta vez de la imagen  $S$  sobre el mismo modelo de color. El resultado de este proceso se muestra en la imagen de la figura B.5(b).

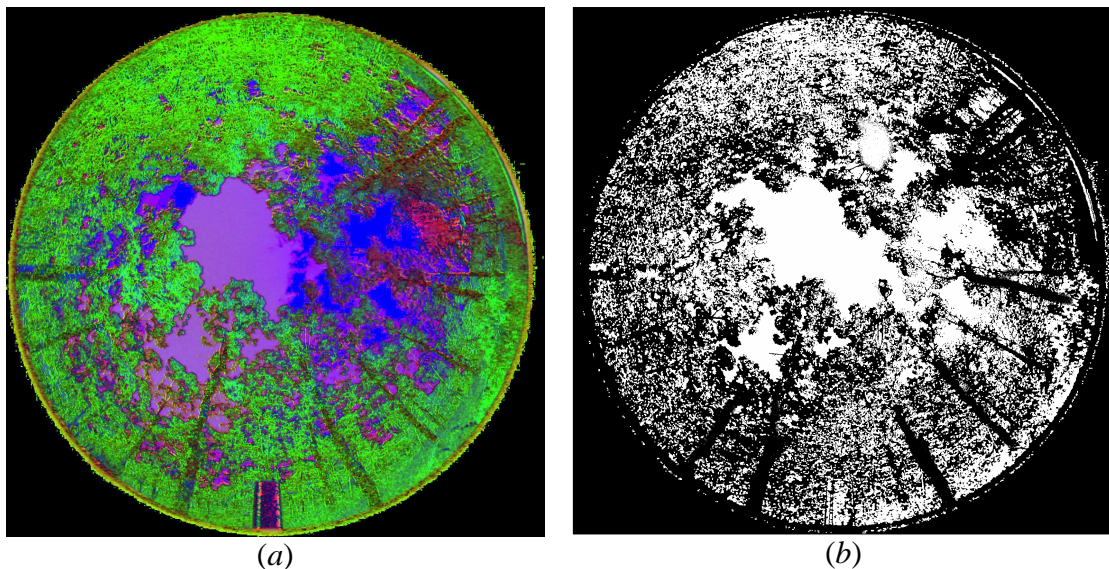


Figura B.5 (a) Imagen de la figura B.1(a) en el modelo de color  $HSI$ ; (b) segmentación mediante la aplicación del método de Otsu.



## Bibliografía

- Abraham, S., Förstner, W. (2005). Fish-eye-stereo calibration and epipolar rectification, *Photogrammetry and Remote Sensing*, 59, 278-288.
- Aiyer, A., Pyun, K., Huang, Y., O'Brien, D.B., Gray, R.M. (2005). Lloyd Clustering of Gauss mixture models for image compression and classification, *Signal Processing: Image Communication*, 20, 459-485.
- Allili, M.S., Ziou, D. (2007). Globally adaptive region information for automatic color texture image segmentation, *Pattern Recognition Letters*, 28, 1946-1956.
- Andrey, P., Tarroux, P. (1998). Unsupervised Segmentation of Markov Random Field Modeled Textured Images Using Selectionist Relaxation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 252-262.
- Ansar, A., Castano, A., Matthies, L. (2004). Enhanced real-time stereo using bilateral filtering, in: *2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, 455-462, Thessaloniki, Greece.
- Ansari, M.E., Masmoudi, L., Bensrhair, A. (2007). A new regions matching for color stereo images, *Pattern Recognition Letters*, 28, 1679-1687.
- Arivazhagan, S., Ganesan, L. (2003). Texture classification using wavelet transform, *Pattern Recognition Letters*, 24, 1513-1521.
- Asuncion, A., Newman, D.J. (2010). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science, <http://www.ics.uci.edu/~mllearn/MLRepository.html> (disponible on-line).
- Atkinson, P.M., Tatnall, A.R.L. (1997). Neural Networks in Remote Sensing, *Int. Journal of Remote Sensing*, 18, 699-709.
- Ayache, N. (1991). *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception*, MIT Press, Cambridge, MA.

- Bakstein, H., Pajdla, T. (2002). Panoramic Mosaicing with a 180° Field of View Lens, in: Proceedings of the Omnidirectional Vision Workshop, pp. 60-68. <ftp://cmp.felk.cvut.cz/pub/cmp/articles/bakstein/Bakstein-Pajdla-Omnivis2002.pdf> (disponible on-line).
- Balasko, B., Abonyi, J., Feil, B. (2008). Fuzzy Clustering and Data Analysis Toolbox for Use with Matlab, Veszprem University, Hungary. <http://www.fmt.vein.hu/softcomp/fclusttoolbox/FuzzyClusteringToolbox.pdf> (disponible on-line).
- Bandzi, P., Oravec, M., Pavlovicova, J. (2007). New Statistics for Texture Classification Based on Gabor Filters, in: Proceedings of Czech and Slovak Technical Universities, Radioengineering, Volume. 16, No. 3, pp. 133-137.
- Banno, A., Ikeuchi, K. (2009). Disparity Map Refinement and 3D Surface Smoothing via Directed Anisotropic Diffusion, IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pp. 1870-1877.
- Barker, S.A., Rayner, P.J.W. (2000). Unsupervised image segmentation using Markov random field models, Pattern Recognition, 33, 587-602.
- Barnard, S., Fishler, M. (1982). Computational Stereo, ACM Computing Surveys, 14, 553-572.
- Barnea, D.I., Silverman, H.F. (1972). A Class of Algorithms for Fast Digital Image Registration, IEEE Trans. Computers, 21, 179-186.
- Barnsley, M.J., Barr, S.L. (1996). Inferring urban land use from satellite sensor images using kernel-based spatial reclassification, Photogrammetric Engineering and Remote Sensing, 62(8), 949-958.
- Benediktsson, J.A., Swain, P.H., Ersoy, O.K. (1990). Neural network approaches versus statistical methods in classification of multisource remote sensing data, IEEE Trans. Geosci. Remote Sens., 28(4), 540-552.
- Bezdek, J.C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms, Kluwer, Plenum Press, New York.
- Borges, J.S., Bioucas-Dias, J.M., Marça, A.R.S. (2007). Bayesian Hyperspectral Image Segmentation with Discriminative Class Learning, in: J. Martí, J.M. Benedí, A.M. Mendonça and J. Serrat (Eds.), Pattern Recognition and Image Analysis, Springer-Verlag, Berlin, LNCS 4477, Part I, pp. 22-29.
- Brodatz, P. (1966). Textures: A photographic album for artists and designers, Dover Publications, New York.
- Broek, E.L., Rikxoort, E.M. (2004). Evaluation of color representation for texture analysis, in: Proceedings of the 16th Belgium-Netherlands Artificial Intelligence Conference, R. Verbrugge, N. Taatgen, and L. R. B. Schomaker, (Eds.), Groningen - The Netherlands, 35-42.
- Campadelli, P., Medici, D., Schettini, R. (1997). Color image segmentation using Hopfield Networks, Image and Vision Computing, 15, 161-166.
- Camps-Valls, G., Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens., 43(6), 1351-1362.
- Camps-Valls, G., Gómez-Chova, L., Calpe, J., Soria, E., Martín, J. D., Alonso, L., Moreno, J. (2004). Robust support vector method for hyperspectral data

- classification and knowledge discovery, *IEEE Trans. Geoscience. Remote Sensing*, 42(7), 1530–1542.
- Cariou, C., Chehdi, K. (2008). Unsupervised texture segmentation/classification using 2-D autoregressive modelling and the stochastic expectation-maximization algorithm, *Pattern Recognition Letters*, 29, 905-917.
- Chamorro-Martínez, J., Galán-Perales, E., Prados-Suárez B., Soto-Hidalgo, J.M. (2007). Perceptually-Based Functions for Coarseness Textural Feature Representation, in: J. Martí, J.M. Benedí. A.M. Mendonça and J. Serrat (Eds.), *Pattern Recognition and Image Analysis*, Springer-Verlag, Berlin, LNCS 4477, Part I, 578-586.
- Chan, J.C.W., Laporte, N., Defries, R.S. (2003). Texture Classification of logged forest in tropical Africa using machine-learning algorithms, *Int. J. Remote Sensing*, 24(6), 1401-1407.
- Chang, M.M., Sezan, M.I., Tekalp, A.M. (1994). Adaptive Bayesian segmentation of color images, *Journal Electronic Imaging*, 3, 404-414.
- Chang, T., Kuo, C.C.J. (1993). Texture analysis and classification with tree structured wavelet transform, *IEEE Trans. Image Processing*, 2(10), 429-441.
- Chanussot, J., Benediktsson, J.A., Fauvel, M. (2006). Classification of Remote Sensing Images from Urban Areas using a Fuzzy Possibilistic Model, *IEEE Geoscience and Remote Sensing Letters*, 3(1), 40-1407.
- Chehata, N., Jung, F., Deseilligny, M.P., Stamon, G. (2003). A Region-Based Matching Approach for 3D-Roof Reconstruction from HR Satellite Stereo Pairs, in: *Proc. VIIth Digital Image Computing: Techniques and Applications*, Sun C., Talbot H., Ourselin S., Adriaansen T., Eds., Sydney, Australia, pp. 889-898.
- Chen, J., Pappas, T.N., Mojsilovic, A., Rogowitz, B.E. (2005). Adaptive Perceptual Color-Texture Image Segmentation, *IEEE Trans. Image Processing*, 14(10), 1524-1536.
- Chen, C.T. (2000). Extensions of the TOPSIS for group decision-making under fuzzy environment, *Fuzzy Sets and Systems*, 114, 1-9.
- Cheng, H.D., Jiang, X.H., Sun, Y., Wang, J. (2001). Color image segmentation: advances and prospects, *Pattern Recognition*, 34 (12), 2259–2281.
- Clark, A.A., Thomas, B.T., Campbell, N.W., Greenway, P. (1999). Texture deconvolution for the Fourier-based analysis of non-rectangular regions. *BMVC*, 193-202.
- Cochran, S.D., Medioni, G. (1992). 3-D Surface Description from binocular stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10), 981-994.
- Cohen, M.A., Grossberg, S.G. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks, *IEEE Trans. Syst. Man Cybern.*, 13, 815-826.
- Comaniciu, D., Meer, P. (2002). Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 603–619.

- Comaniciu, D., Meer, P. (1997). Robust analysis of feature spaces: color image segmentation, in: Proc. IEEE Conf. Computer Vision and Pattern Recognition, San Juan, PR, 750-755.
- Debeir, O., van den Steen, I., Latinne, P., van Ham, P., Wolff, E. (2002). Textural and contextual land-cover classification using single and multiple classifier systems, *Photogramm. Eng. Remote Sens.*, 68(6), 597-605.
- Dempster, A.P. (1968). A generalization of Bayesian inference, *Journal of the Royal Statistical Society, B* 30, 205-247.
- Deng, H., Clausi, D.A. (2005a). Unsupervised segmentation of synthetic aperture radar ice imagery using a novel Markov random field model, *IEEE Trans. Geosci. Remote Sensing*, 43(3), 528-538.
- Deng, H., Clausi, D.A. (2005b). Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov model, *Lecture Notes in computer Science*, Calgari, Italy, 1857, 1-15.
- Deng, H., Clausi, D.A. (2004). Unsupervised image segmentation using a simple MRF model with a new implementation scheme, *Pattern Recognition*, 37(12), 2323-2335.
- Dimap (2010). <http://www.dimap.es/> (disponible on-line).
- Du, Y., Chang, C., Thouin, P.D. (2004). Unsupervised approach to color video thresholding, *Optical Engineering*, 43(2), 282-289.
- Duda, R.O., Hart, P.E., Stork, D.S. (2000). *Pattern Classification*, 2nd Ed., Wiley.
- Dunn, D., Higgins, W.E. (1995). Optimal Gabor filters for texture segmentation, *IEEE Trans. Image Processing*, 4(7), 947-964.
- Elias, R. (2007). Sparse view stereo matching, *Pattern Recognition Letters*, 28, 1667-1678.
- Escalera, A. (2001). *Visión por Computador: Fundamentos y Métodos*, Prentice-Hall, Madrid.
- Ferreiro-Armán, M., Bandeira, L.P.C., Martín-Herrero, J., Pina, P. (2007). Classifiers for vegetation and forest mapping with low resolution multispectral imagery, in: J. Martí, J.M. Benedí, A.M. Mendonça and J. Serrat (Eds.), *Pattern Recognition and Image Analysis*, Springer-Verlag, Berlin, LNCS 4477, Part I, 177-184.
- Frate, F., Del, Pacifici, F., Schiavon, G., Solimini, C. (2007). Use of Neural Networks for Automatic Classification from High-Resolution Images, *IEEE Trans. Geoscience and Remote Sensing*, 45(4), 800-809.
- Frizzelle, B.G., Moody, A. (2001). Mapping continuous distributions of land cover: A comparison of maximum-likelihood estimation and artificial neural networks, *Photogramm. Eng. Remote Sens.*, 67(6), 693-705.
- Geman, S., Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Analysis Machine Intelligence*, 6, 721-741.

- Giacinto, G., Roli, F., Bruzzone, L. (2000). Combination of neural and statistical algorithms for supervised classification of remote-sensing image, *Pattern Recognition Letters*, 21(5), 385-397.
- Gong, P., Howarth, P.J. (1990). The use of structural information for improving land-cover classification accuracies at the rural-urban fringe, *Photogramm. Eng. Remote Sens.*, 56(1), 67-73.
- Gonzalez, R.C., Woods, R.E. (2008). *Digital Image Processing*, 3rd Ed., Prentice-Hall, New Jersey, USA.
- Gregoire, T.G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference, *Canadian Journal of Forest Research*, 28, 1429-1447.
- Grimson, W.E.L. (1985). Computational experiments with a feature-based stereo algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7, 17-34.
- Guijarro, M. (2009). *Combinación de clasificadores para identificación de texturas en imágenes naturales: nuevas estrategias locales y globales*, Tesis doctoral, Departamento de Ingeniería del Software e Inteligencia Artificial, Facultad de Informática, Universidad Complutense de Madrid.
- Guijarro, M., Pajares, G. (2009). On combining classifiers through a fuzzy Multicriteria Decision Making Approach: applied to natural textured images, *Expert Systems with Applications*, vol. 36, no. 3P2, 7662-7669.
- Guijarro, M., Pajares, G., Herrera, P.J. (2009). Image-Based Airborne Sensors: A Combined Approach for Spectral Signatures Classification through Deterministic Simulated Annealing, *Sensors*, vol. 9, issue 9, 7132-7149.
- Guijarro, M., Pajares, G., Herrera, P.J. (2008). On Combining Classifiers by Relaxation for Natural Textures in Images, in: E. Corchado, A. Abraham, and W. Pedrycz (Eds.), *HAIS 2008, LNAI 5271*, Springer-Verlag Berlin Heidelberg, pp. 345-352.
- Haapanen, R., Ek, A.R., Bauer, M.E., Findley, A.O. (2004). Delineation of forest/nonforest land use classes using nearest neighbor methods, *Remote Sensing Environment*, 89(3), 265-271.
- Hanmandlu, M., Madasu, V.K., Vasikarla, S. (2004). A Fuzzy Approach to Texture Segmentation, in: *Proc. of the IEEE International Conference on Information Technology: Coding and Computing (ITCC'04)*, The Orleans, Las Vegas, Nevada, USA, 636-642.
- Hansen, M., Dubayah, R., Defries, R. (1996). Classification trees: an alternative to traditional land cover classifiers, *Int. J. Remote Sensing*, 17(5), 1075-1081.
- Haralick, R.M., Shapiro, L.G. (1992). *Computer and Robot Vision*, vol. I y II, Addison-Wesley: Reading, MA.
- Haykin, S. (1994). *Neural Networks: A comprehensive Foundation*, McMillan College Publishing Co., New York.
- Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J., Cruz, J.M. (2009a). Choquet Fuzzy Integral applied to stereovision matching for fish-eye lenses in forest analysis, in: W. Yu and E.N. Sanchez (Eds.), *Advances in Computational Intell.*, AISC 61, Springer-Verlag Berlin Heidelberg, pp. 179-187.



- Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J., Cruz, J.M. (2009*b*). Combination of attributes in stereovision matching for fish-eye lenses in forest analysis, in: J. Blanc-Talon et al. (Eds.), ACIVS 2009, LNCS 5807, Springer-Verlag Berlin Heidelberg, pp. 277-287.
- Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J., Cruz, J.M. (2009*c*). Fuzzy Multi-Criteria Decision Making in Stereovision Matching for Fish-Eye Lenses in Forest Analysis, in: H. Yin and E. Corchado (Eds.), IDEAL 2009, LNCS 5788, Springer-Verlag Berlin Heidelberg, pp. 325-332.
- Herrera, P.J., Pajares, G., Guijarro, M., Ruz, J.J., Cruz, J.M., Montes, F., (2009*d*). A Featured-Based Strategy for Stereovision Matching in Sensors with Fish-Eye Lenses for Forest Environments, *Sensors*, 9(12), 9468-9492.
- Hopfield, J.J., Tank, D.W. (1986). Computing with neural circuits: a model, *Science*, 233, 625-633.
- Hopfield, J.J., Tank, D.W. (1985). Neural computation of decisions in optimization problems, *Biol. Cyber.*, 52, 141-152.
- Howard, I.P., Rogers, B.J. (1995). *Binocular Vision and Stereopsis*, Oxford University Press, New York.
- Hu, Q., Yang, Z. (2008). Stereo Matching Based on Local Invariant Region Identification, in: *The International Symposium on Computer Science and Computational Technology*, Volume 2, pp.690-693.
- Huang, C.L. (1999). Pattern image segmentation using modified Hopfield model, *Pattern Recognition Letters*, 13, 345-353.
- ISCAR. (2006). Grupo Ingeniería de Sistemas, Control, Automatización y Robótica. [www.dacya.ucm.es/area-isa](http://www.dacya.ucm.es/area-isa)
- Jiji, G.W., Ganesan, L. (2008). A new approach for unsupervised segmentation, *Applied Soft Computing Journal*, doi:10.1016/j.asoc.2008.06.002.
- Joya, G., Atencia, M.A., Sandoval, F. (2002). Hopfield neural networks for optimization: study of the different dynamics, *Neurocomputing*, 43, 219-237.
- Kaick, O.V., Mori, G. (2006). Automatic Classification of Outdoor Images by Region Matching, in: *The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, pp.9.
- Kasetkasem, T., Varshney, P.K. (2002). An image change detection algorithm based on Markov Random field models, *IEEE Trans. Geoscience Remote Sensing*, 40(8), 1815-1823.
- Kim, J., Fisher III, J.W., Yezzi, A., Çetin, M., Willsky, A.S. (2006). A non parametric statistical method for image segmentation using information theory and curve evolution, *IEEE Trans. Image Processing*, 14(10), 1486-1502.
- Klaus, A., Sormann, M., Karner, K. (2006). Segmented-Based Stereo Matching Using Belief Propagation and Self-Adapting Dissimilarity Measure, In: *Proc. of 18th Int. Conference on Pattern Recognition*, vol. 3, pp. 15-18.
- Koffka, K. (1935). *Principles of Gestalt Psychology*, New York: Harcourt.
- Koschan, A., Abidi, M. (2008). *Digital Color Image Processing*, Wiley, Chichester.

- Kosko, B. (1992). *Neural Networks and Fuzzy Systems: a dynamical systems approach to machine intelligence*, Prentice-Hall, NJ.
- Kosko, B. (1986). Fuzzy Cognitive Maps, *Int. J. Man Machine Studies*, 24, 65-75.
- Krishnamachari, S., Chellappa, R. (1997). Multiresolution Gauss-Markov random field models for texture segmentation, *IEEE Transactions on Image Processing*, 6(2), 251-267.
- Krishnapuram, R., Freg, C.P. (1992). Fitting an unknown number of lines and planes to image data through compatible cluster merging, *Pattern Recognition*, 25, 385-400.
- Krotkov, E., Henriksen, K., Kories, R. (1990). Stereo ranging with verging cameras, *IEEE Trans. Pattern Anal. Machine Intell.*, 12(12), 1200-1205.
- Krotkov, E. (1989). *Active Computer Vision by Cooperative Focus and Stereo*, Springer-Verlag, Berlín.
- Krotkov, E. (1987). Focusing, *International Journal of Computer Vision*, 1(3), 223-237.
- Kumar, A.S., Majumder, K.L. (2001). Information fusion in tree classifiers, *Int. J. Remote Sensing*, 22(5), 861-869.
- Kumar, S., Ghosh, J., Crawford, M.M. (2001). Best-bases feature extraction for pairwise classification of hyperspectral data, *IEEE Trans. Geoscience and Remote Sensing*, 39(7), 1368-1379.
- Kuncheva, L. (2004). *Combining Pattern Classifiers: Methods and Algorithms*, Wiley, Chichester.
- Kupinski, M.A., Gigger, M.L. (1998). Automated seeded lesion segmentation on digital mammograms, *IEEE Trans. Medical Imaging*, 17, 510-517.
- Laarhoven, P.M.J. van, Aarts, E.H.L. (1989). *Simulated Annealing: Theory and Applications*, Kluwer Academic, Holland.
- Lankton, S. (2010), <http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/> (disponible on-line).
- Lee, D.L., Chuang, T.C. (2005). Designing asymmetric Hopfield-type associative memory with higher order Hamming stability, *IEEE Trans. Neural Networks*, 16(6), 1464-1476.
- Lew, M.S., Huang, T.S., Wong, K. (1994). Learning and feature selection in stereo matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 869-881.
- Li, H., Liu, G., Zhang, Z. (2006). A new texture generation model based on Pseudo-DCT coefficients, *IEEE Trans. Image Processing*, 15(5), 1300-1312.
- Li, J., Najmi, A., Gray, R.M. (2000). Image classification by a two-dimensional hidden Markov model, *IEEE Trans. Signal Processing*, 48(2), 517-533.
- Liew, A.W.C., Leung, S.H., Lau, W.H. (2003). Segmentation of color images by spatial fuzzy clustering, *IEEE Trans. Fuzzy Systems*, 11, 542-549.
- Lillo, A., Motta, G., Storer, J. A. (2007). Supervised segmentation based on texture signatures extracted in the Frequency Domain, in: J. Martí, J.M. Benedí. A.M.

- Mendoça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Springer-Verlag, Berlin, LNCS 4477, Part I, 89-96.
- Lim, Y.W., Lee, S.U. (1990). On the color image segmentation algorithm based on the thresholding and the fuzzy c-means technique, *Pattern Recognition*, 23(9), 935-952.
- Lipton, L. (1997). Stereo-vision formats for video and computers graphics, StereoGraphics Corporation.
- Littmann, E., Ritter, H. (1997). Adaptive color segmentation -A comparison of neural and statistical methods, *IEEE Trans. Neural Networks*, 8(1), 175-185.
- Liu, F., Picard, R.W. (1996). Periodicity, directionality and randomness: Wold features for image modelling and retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7), 722-733.
- Lo, Y.S., Pei, S.C. (1999). Color image segmentation using local histogram and self-organizing of Kohonen feature map, in: *Proc. Int. Conf. on Image Processing*, Kobe, Japan, 232-239.
- López, M.A., Pla, F. (2000). Dealing with Segmentation Errors in Region-based Stereo Matching, *Pattern Recognition*, 8(33), pp. 1325-1338.
- Luo, J., Gray, R.T., Lee, H.C. (1998). Incorporation of derivative priors in adaptive Bayesian color image segmentation, in: *Proc. Int. Conf. Image Processing*, Chicago, 3, 780-784.
- Mandallaz, D., Ye, R. (1999). Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance, *Canadian Journal of Forest Research*, 29, 1691-1708.
- Manjunath, B.S., Ma, W.Y. (1996). Texture features for browsing and retrieval of image data, *IEEE Trans. Pattern Analysis and Machine Intell.*, 18(8), 837-842.
- Manjunath, B.S., Symchony, T., Chellappa, R. (1990). Stochastic and deterministic networks for texture segmentation, *IEEE Trans. Acoustic, Speech, Signal Processing*, 38(6), 1039-1049.
- Marapane, S.B., Trivedi, M.M. (1989). Region-based stereo analysis for robotic applications, *IEEE Transactions on Systems, Man and Cybernetics* 19(6), 1447-1464.
- Mas, J.F., Flores, J.J. (2008). The application of artificial neural networks to the analysis of remotely sensed data, *Int. J. Remote Sensing*, 29(3), 617-663.
- MATLAB. (2010). The Mathworks. <http://www.mathworks.com/products/matlab/> (disponible on-line).
- Matthies, L., Maimone, M., Johnson, A., Cheng, Y., Willson, R., Villalpando, C., Goldberg, S., Huertas, A. (2007). Computer Vision on Mars, *International Journal of Computer Vision*, 75(1), 67-92.
- McKinnon, B., Baltes, J. (2004). Practical Region-Based Matching for Stereo Vision, in: *10th International Workshop on Combinatorial Image Analysis (IWCIA'04)*, Klette, R., Zunic, J., Eds., Springer, LNCS 3322, pp. 726-738.
- Medioni, G., Nevatia, R. (1985). Segment Based Stereo Matching, *Computer Vision, Graphics and Image Processing*, 31, 2-18.

- Melas, D.E., Wilson, S.P. (2002). Double Markov random fields and Bayesian image segmentation, *IEEE Transactions on Signal Processing*, 50 (2), 357-365.
- Melgani, F., Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines, *IEEE Trans. Geosci. Remote Sensing*, 42(8), 1178-1790.
- Meyer, P., Roubens, M. (2006). On the use of the Choquet integral with fuzzy numbers in multiple criteria decision support, *Fuzzy Sets and Systems*, 157, 927-938.
- Miao, Y., Liu, Z.Q. (2000). On Causal Inference in Fuzzy Cognitive Maps, *IEEE Trans. Fuzzy Systems* 8(1), 107-119.
- Middlebury. (2010). <http://vision.middlebury.edu/stereo/> (disponible on-line).
- Mohaddamzadeh A., Bourbakis, N. (1997). A fuzzy region growing approach for segmentation of color images, *Pattern Recognition*, 30(6), 867-881.
- Montes, F., Ledo, A., Rubio, A., Pita, P., Cañellas, I. (2009). Use of estereoscopic hemispherical images for forest inventories, in: *Proc. Int. Scientific Conference Forest, Wildlife and Wood Sciences for Society development*, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic.
- Montes, F., Hernández, M.J., Cañellas, I. (2005). A geostatistical approach to cork production sampling estimation in *Quercus suber* L. forests, *Canadian Journal of Forest Research*, 35, 2787-2796.
- Montiel, E., Aguado, A.S., Nixon, M.S. (2005). Texture classification via conditional histograms, *Pattern Recognition Letters*, 26, 1740-1751.
- Müezzinoğlu, M.K., Güzelis, C., Zurada, J.M. (2005). An energy function-based design method for discrete Hopfield associative memory with attractive fixed points, *IEEE Trans. Neural Networks*, 16(2), 370-378.
- Muñoz-Marí, J., Bruzzone, L., Camps-Valls, G. (2007). A Support Vector Domain Description Approach to Supervised Classification of Remote Sensing Images, *IEEE Trans. Geosci. Remote Sens.*, 45(8), 2683-2692.
- Newsam, S., Wang, L. Bhagavathy, S., Manjunath, B.S. (2004). Using Texture to analyze and Manage Large Collections of Remote Sensed image and Video Data, *Applied Optics*, 43(2), 210-217.
- Ogale, A.S., Aloimonos, Y. (2005). Shape and the Stereo Correspondence Problem, *International Journal of Computer Vision*, 65(3), 147-162.
- Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp. 62-66.
- Özden, M., Polat, E. (2007). A color image segmentation approach for content-based image retrieval, *Pattern Recognition*, 40(4), 1318-132.
- Pajares, G., Guijarro, M., Ribeiro, A. (2010). A Hopfield Neural Network for combining classifiers applied to textured images, *Neural Networks*, 23(1), 144-153.
- Pajares, G., Guijarro, M., Herrera, P.J., Ribeiro, A. (2009). Combining Classifiers through Fuzzy Cognitive Maps in natural images, *IET Computer Vision*, vol. 3, issue 3, 112-123.

- Pajares, G., Cruz, J.M. (2007a). *Visión por Computador: Imágenes digitales y aplicaciones*, 2ª ed., RA-MA, Madrid.
- Pajares, G., Cruz, J.M. (2007b). *Ejercicios resueltos de Visión por Computador*, RA-MA, Madrid.
- Pajares, G., Cruz, J.M. (2006). Fuzzy Cognitive Maps for stereovision matching, *Pattern Recognition*, 39, 2101–2114.
- Pajares, G., Cruz, J.M. (2002). Clasificación de Texturas Naturales mediante K-Means, *Revista Electrónica de Visión por Computador*, (<http://revc.uab.es/revista/06/>, ISSN 1575-5258), no. 6, 1-18, Ed. Centre de Visió per Computador, Universitat Autònoma, Barcelona.
- Pajares, G., Cruz, J.M., Moreno, V. (2002). Clasificación de texturas naturales mediante agrupamiento borroso, *Ingeniería Civil*, Centro de Estudios y Experimentación de Obras Públicas (CEDEX), Ministerio de Fomento, 127, 83-89.
- Pajares, G., Cruz, J.M., Moreno, V. (2001). Clasificación de texturas mediante redes neuronales, *Ingeniería Civil*, Centro de Estudios y Experimentación de Obras Públicas (CEDEX), Ministerio de Fomento, 123, 61-69.
- Pajares, G., Cruz, J.M., Aranda, J. (1998). Relaxation by Hopfield Network in Stereo Image Matching, *Pattern Recognition*, 31(5), 561 – 574.
- Palm, C., Lehmann, T.M. (2002). Classification of color textures by Gabor filtering, *Machine Graphics and Vision*, 11(2/3), 195-219.
- Paola, J.D., Schowengerdt, R.A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery, *Int. J. Remote Sensing*, 16(16), 3033-3058.
- Pappas, T.N. (1992). An adaptive clustering algorithm for image segmentation, *IEEE Trans. Signal Processing*, 40(4), 901-914.
- Pardé, J., Bouchon, J. (1987). *Dendrométrie*; l'École National du Génie Rural des Eaux et des Forêts, Nancy, France.
- Permuter, H., Francos, J., Jermyn, I. (2006). A study of Gaussian mixture models of color and texture features for image classification and segmentation, *Pattern Recognition*, 39, 695-706.
- Pi, M.H., Tong, C.S., Choy, S.K., Zhang, H. (2006). A fast and effective model for wavelet subband histograms and its applications in texture image retrieval, *IEEE Trans. Image Processing*, 15(10), 3078-3088.
- Pita, P.A. (1973). *El Inventario en la Ordenación de Montes*, INIA, Ministerio de Agricultura, Pesca y Alimentación, Madrid, Spain.
- Pollard, S.B., Mayhew, J.E.W., Frisby, J.P. (1981). PMF: A stereo correspondence algorithm using a disparity gradient limit, *Perception*, 14, 449-470.
- Porat, M., Zeevi, Y.Y. (1989). Localized texture processing in vision: analysis and synthesis in the Gaborian space, *IEEE Transactions on Biomedical Engineering*, 36(1), 115-129.
- Pratt, W.K. (1991). *Digital Image Processing*, John Wiley and Sons, NY.

- Precioso, F., Barlaud, M., Blu, T., Unser, M. (2005). Robust real time segmentation of images and videos using a smooth-spline snake-based algorithm, *IEEE Trans. Image Processing*, 14(7), 910-924.
- Premaratne, P., Safaei, F. (2008). Feature based Stereo correspondence using Moment Invariant, in: *Proc. 4<sup>th</sup> Int. Conf. Information and Automation for Sustainability (ICIAFS'08)*, pp.104-108.
- Puig, D., García, M.A. (2006). Automatic texture feature selection for image pixel classification, *Pattern Recognition*, 39(11), 1996-2009.
- Pyun, K., Lim, J. Won, C.S., Gray, R.M. (2007). Image Segmentation using Hidden Markov Gauss mixture models, *IEEE Trans. Image Processing*, 16(7), 1902-1911.
- Pyun, K., Won, C.S., Lim, J., Gray, R.M. (2002). Texture classification based on multiple Gauss mixture vector quantizer, in: *Proc. of ICME*, 501-504.
- Qiao, H., Peng, J., Xu, Z.B. (2001). Nonlinear Measures: A new approach to exponential stability analysis for Hopfield-type neural networks, *IEEE Trans. Neural Networks*, 12(2), 360-370.
- Randen, T., Husøy, J.H. (1999). Filtering for Texture Classification: A Comparative Study, *IEEE Trans. Pattern Analysis Machine Intell.*, 21(4), 291-310.
- Reid, I.D., Beardsley, P.A. (1996). Self-alignment of a binocular robot, *Image Vision computing*, 14, 635-640.
- Renninger, L.W., Malik, J. (2004). When is scene recognition, just texture recognition?, *Vision Research*, 44, 2301-2311.
- Robinson, G.S. (1977). Color edge detection, *Optical Engineering*, 16(5), 479-484.
- Rohith, M.V., Somanath, G., Kambhamettu, C., Geiger, C. (2008). Towards estimation of dense disparities from stereo images containing large textureless regions, in: *ICPR 2008, 19th International Conference on Pattern Recognition*, 1-5, Tampa, FL.
- Rovira-Más, F., Zhang, Q., Reid, F. (2008). Stereo vision three-dimensional terrain maps for precision agriculture, *Computers and Electronics in Agriculture*, 60, 133-143.
- Rud, R., Shoshany, M., Alchanatis, V., Cohen, Y. (2006). Application of spectral features' ratios for improving classification in partially calibrated hyperspectral imagery: a case study of separating Mediterranean vegetation species, *Journal Real-Time Image Processing*, 1, 143-152.
- Ruichek, Y., Postaire, J.G. (1996). A neural matching algorithm for 3-D reconstruction from stereo pairs of linear images, *Pattern Recognition Letters*, 17, 387-398.
- Sarkar, A., Biswas, M.K., Sharma, K.M.S. (2000). A simple unsupervised MRF model based image segmentation approach, *IEEE Transactions on Image Processing*, 9(5), 801-812.
- Scaramuzza, D., Cribblez, N., Martinelli, A., Siegwart, R. (2008). Robust Feature Extraction and Matching for Omnidirectional Images, *Field and Service Robotics*, Laugier, C., Siegwart, R., Eds., Springer, Berlin, Germany, Volume 42, pp. 71-81.



- Scharstein, D., Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *Int. J. Computer Vision*, vol. 47(1-3), pp. 7–42.
- Schwalbe, E. (2005). Geometric Modelling and Calibration of Fisheye Lens Camera Systems, in: *Proc. 2nd Panoramic Photogrammetry Workshop*, *Int. Archives of Photogrammetry and Remote Sensing*, Vol. 36, Part 5/W8.
- Seong, J.C., Usery, E.L. (2001). Fuzzy image classification for continental scale multitemporal NDVI series images using invariant pixels and an image stratification method, *Photogrammetric Engineering and Remote Sensing*, 67(3), 287-294.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press, Princeton.
- Shah, S., Aggarwal, J.K. (1997). Mobile robot navigation and scene modeling using stereo fish-eye lens system, *Machine Vision and Applications*, 10, 159–173.
- SIC, Satellite Image Corporation (2010). <http://www.satimagingcorp.com/>.
- Singh, S., Haddon, J., Markou, M. (2001). Nearest-neighbour classifiers in natural scene analysis, *Pattern Recognition*, 34, 1601-1612.
- Smith, G., Burns, I. (1997). Measuring texture classification algorithms, *Pattern Recognition Letters*, 18, 1495-1501.
- Starink, J.P., Backer, E. (1995). Finding Point Correspondences Using Simulated Annealing, *Pattern Recognition*, 28(2), 231-240.
- Suárez, A. (2009). Análisis de métodos de procesamiento de imágenes estereoscópicas forestales. Trabajo Fin de Máster, Facultad de Informática, Universidad Complutense.
- Tang, L., Wu, C., Chen, Z. (2002). Image dense matching based on region growth with adaptive window, *Pattern Recognition Letters*, vol. 23, pp. 1169–1178.
- Tao, Y., Lam, E.C.M., Tang, Y.Y. (2000). Extraction of fractal feature for pattern recognition, *Proceedings International Conference on Pattern Recognition*, vol. 2. IAPR, Barcelona, Spain, pp. 527–530.
- Tax, D.M.J., Breukelen, van M., Duin, R.P.W., Kittler, J. (2000). Combining multiple classifiers by averaging or by multiplying?, *Pattern Recognition*, 33, 1475-1485.
- Tkalcic, M., Tasic, J.F. (2003). Colour spaces-perceptual, historical and applicational background, in: *EUROCON 2003, Computer as a Tool, The IEEE Region 8*, vol. 1, pp. 304-308.
- Tottrup, C. (2007). Forest and Land Cover Mapping in a Tropical Highland Region, *Photogrammetric Engineering and Remote Sensing*, 73(9), 1057-1066.
- Trauwaert, E., Kaufman, L., Rousseeuw, P. (1991). Fuzzy Clustering Algorithms Based on the Maximum Likelihood Principle, *Fuzzy Sets and Systems*, 42, 213-227.
- Trias-Sanz, R., Stamon, G., Louchet, J. (2008). Using colour, texture, and hierarchical segmentation for high-resolution remote sensing, *ISPRS Journal of Photogrammetry & Remote Sensing*, 63, 156-168.
- Truevision3D. (2010). <http://www.truevision3d.com/> (disponible on-line).

- Tsardias, A.K., Margaritis, K.G. (1999). An experimental study of the dynamics of the certainty neuron fuzzy cognitive maps, *Neurocomputing*, 24, 95-116.
- Tsardias, A.K., Margaritis, K.G. (1997). Cognitive Mapping and Certainty Neuron Fuzzy Cognitive Maps, *Information Sciences* 101, 109-130.
- Tso, B.C.K., Mather, P.M. (1999). Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields, *IEEE Trans. Remote Sensing*, 37(3), 1255-1260.
- Unser, M. (1995). Texture classification and segmentation using wavelet frames, *IEEE Trans. Image Processing*, 4(11), 1549-1560.
- Valdovinos, R.M., Sánchez, J.S., Barandela, R. (2005). Dynamic and Static weighting in classifier fusion, in: J.S. Marques, N. Pérez de la Blanca and P. Pina (Eds.), *Pattern Recognition and Image Analysis*, Springer-Verlag, Berlin, LNCS 3523, pp. 59-66.
- Vazquez, E., Baldrich, R. Vazquez, J., Vanrell, M. (2007). Topological Histogram Reduction towards colour segmentation, in: J. Martí, J.M. Benedí. A.M. Mendonça and J. Serrat (Eds.), *Pattern Recognition and Image Analysis*, Springer-Verlag, Berlin, LNCS 4477, Part I, 55-62.
- Vergara, M.L. (1988). *Manual de Fotogeología*, CIEMAT.
- VisTex (2010). <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html> (disponible on-line).
- Wagner, T. (1999). Texture Analysis, in: B. Jahne, H. Haussecker, P. Geisser (Eds.), *Handbook of Computer Vision and Applications*, Academic Press, San Diego, cap. 2 (Signal Processing and Pattern Recognition).
- Wan, T., Canagarajah, N., Achim, A. (2007). Multiscale Color-Texture image segmentation with adaptive Region Merging, In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP08)*, 1, 1-1213-1216.
- Wang, Z.F., Zheng, Z.G. (2008). A region based stereo matching algorithm using cooperative optimization, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, pp. 1-8.
- Wang, Z., Boesch, R. (2007). Color- and Texture-Based Image Segmentation for Improved Forest Delineation, *IEEE Trans. Geoscience and Remote Sensing*, 45(10), 3055-3062.
- Wang, W., Fenton, N. (2006). Risk and confidence analysis for fuzzy multi criteria decision making, *Knowledge Based Systems*, 19, 430-437.
- Wang, D. (2005). The time dimension for scene analysis, *IEEE Trans. Neural Networks*, 16(6), 1401-1426.
- Wangenheim, A., Bertoldi, R.F., Abdala, D.D., Richter, M.M. (2007). Color image segmentation guided by a color gradient network, *Pattern Recognition Letters*, 28, 1795-1803.
- Wei, Y., Quan, L. (2004). Region-Based Progressive Stereo Matching, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, vol. 1, pp. 106-113.



- Wilkinson, G.G. (2005). Results and implications of a Study of Fiteen Years of Satellite Image Classification Experiments, *IEEE Trans. Geoscience Remote Sensing*, 45(3), 433-440.
- Wulder, M.A., Franklin, S.E. (2003). *Remote Sensing of forest environments: concepts and case studies*, Kluwer Academic Publishers, Boston, USA.
- Xia, Y., Feng, D., Wang, T., Zhao, R., Zhang, Y. (2007). Image segmentation by clustering of spatial patterns, *Pattern Recognition Letters*, 28, 1548-1555.
- Xia, Y., Feng, D., Zhao, R. (2006). Adaptive Segmentation of Textured Images by using the coupled Markov Random Field Model, *IEEE Trans. Image Processing*, 15(11), 3559-3566.
- Xie, J., Tsui, H.T. (2004). Image Segmentation based on maximum-likelihood estimation and optimum entropy-distribution (MLE-OED), *Pattern Recognition Letters*, 25(10), 1133-1141.
- Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multi-criteria decision making, *IEEE Trans. System Man and Cybernetics*, 18(1), 183-190.
- Yang, A.Y., Wright, J., Ma, Y., Sastry, S.S. (2008). Unsupervised segmentation of natural images via lossy data compression, *Computer Vision and Image Understanding*, 110, 212-225.
- Yang, F., Jiang, T. (2003). Pixon-Based Image Segmentation with Markov Random Fields, *IEEE Transactions on Image Processing*, 12(12), 1552-1559.
- Yu, H., Li, M., Zhang, H.J., Feng, J. (2002). Color texture moments for content-based image retrieval, in: *Proc. International Conference on Image Processing*, 3, 24-28.
- Yu, S.S., Tsai, W.H. (1992). Relaxation by the Hopfield Neural Network, *Pattern Recognition*, 25(2), 197-209.
- Zhang, Y., Wang, R. (2004). A combined method for texture analysis and its application, in: F. Bubak, M. (Eds.), *Lecture Notes in Computer Science*, 3036, 413-416.
- Zhang, J., Foody, G.M. (1998). A fuzzy classification of sub-urban land cover from remotely sensed imagery, *Int. J. Remote Sensing*, 19(14), 2721-2738.
- Zhao, H. (2004). Global asymptotic stability of Hopfield neural network involving distributed delays, *Neural Networks*, 17, 47-53.
- Zheng, J.Y. (2000). Virtual Recovery and Exhibition of Heritage, *IEEE Multimedia*, 7(2), 31-34.
- Zimmermann, H.J. (1991). *Fuzzy set theory and its applications*, Kluwer Academic Publishers, Norwell, MA.